Equality and Human Rights Commission
Research report 113

Unconscious bias training:

An assessment of the evidence for effectiveness

Doyin Atewologun, Tinu Cornish and Fatima Tresh

Equality and Human Rights Commission

First published March 2018
ISBN: 978-1-84206-720-8

Post: Research Team
Equality and Human Rights Commission
Arndale House
The Arndale Centre
Manchester M4 3AQ

Email: research@equalityhumanrights.com
Telephone: 0161 829 8500
Website: www.equalityhumanrights.com

You can download a copy of this report as a PDF from our website:
www.equalityhumanrights.com

If you require this publication in an alternative format, please contact the Communications Team to discuss your needs at: correspondence@equalityhumanrights.com

Equality and Human Rights Commission Research Report Series

The Equality and Human Rights Commission Research Report Series publishes research carried out for the Commission by commissioned researchers.

The views expressed in this report are those of the authors and do not necessarily represent the views of the Commission. The Commission is publishing the report as a contribution to discussion and debate.

Please contact the Research Team for further information about other Commission research reports, or visit our website.

Contents

Acknowledgements

The authors (who contributed equally to this report) would like to thank Pete Jones and Rob Briner for their helpful feedback. Thanks also to our partners at the Equality and Human Rights Commission, Kathleen Jameson, Rosie Wallbank and David Perfect. Thank you to the Equality Challenge Unit (ECU) for supporting Tinu Cornish's involvement in the review.

Executive summary

In her 2017 review, 'Race in the Workplace', Baroness McGregor-Smith highlighted the 'structural, historical bias' that prevents ethnic minorities, women, disabled people and others from progressing in their careers. She recommended that the UK Government create a free, online unconscious bias training (UBT) resource to tackle the unconscious bias that she described as 'much more pervasive and more insidious than the overt racism that we associate with the 1970s' (McGregor-Smith, 2017, p.2).

Prior to the McGregor-Smith review, and as a direct consequence of it, an increasing number of organisations in the UK have introduced UBT. This training has been implemented even though some academic research and reports have highlighted the ineffectiveness, and even the negative effects, of UBT. Additionally, there remains much academic debate about the accuracy of the Implicit Association Test (IAT) (a reaction-time measure of how quickly a participant can link positive and negative stimuli to labels such as 'male' or 'female'), which is the most common measure of unconscious bias.

The Equality and Human Rights Commission ('the Commission') seeks to contribute to UK debate and policy on the use of UBT to counter workplace inequalities. This report was commissioned to identify and evaluate available evidence to help determine whether, when and how UBT works. It consisted of a rapid evidence assessment methodology. This required a transparent and systematic approach to the search for evidence and the elimination of studies that did not meet pre-specified minimum quality standards.

The research question that this assessment aimed to address was:

- What is the evidence for the effectiveness of unconscious bias training?

What is unconscious bias training?

Unconscious (or implicit) biases, unlike conscious biases, are the views and opinions that we are unaware of; they are automatically activated and frequently operate

outside conscious awareness and affect our everyday behaviour and decision making. Our unconscious biases are influenced by our background, culture, context and personal experiences.

Primarily, UBT aims to increase awareness of unconscious bias and its impact on people who belong to groups denoted as having 'protected characteristics' under the Equality Act 2010 (age, race, sex, disability, religion or belief, gender reassignment, sexual orientation, marriage and civil partnership, pregnancy and maternity). Other aims are to: reduce implicit/unconscious bias towards members of a group denoted as having a 'protected characteristic'; reduce explicit bias towards members of a group denoted as having a 'protected characteristic'; and change behaviour, in the intended direction, towards equality-related outcomes.

UBT is often delivered online to an individual participant or face-to-face as a workshop to a group of participants. Although each experience is different, most UBT interventions include one or more of the following:

- An unconscious bias 'test' (a reaction-time measure of how quickly a participant can link positive and negative stimuli to labels such as 'male' or 'female'; the most common example is the IAT).
- An unconscious bias 'test' debrief (an explanation of the participants' unconscious bias 'test' results).
- Education on unconscious bias theory.
- Information on the impact of unconscious bias (via statistics/illustrative examples).
- Suggested techniques for either reducing the level of unconscious bias or mitigating the impact of unconscious bias (without altering or reducing the strength of the bias). For example, bias reduction strategies, such as exposing participants to counter-stereotypic exemplars, can reduce the level of unconscious bias; bias mitigation strategies, such as blind review in selection and assessment, can reduce the impact of unconscious bias.

Key findings

Overall, our evaluation of rigorous studies on the effectiveness of UBT indicates a mixed picture and a need for further research to determine the effectiveness of unconscious bias training. We found that:

- UBT is effective for awareness raising by using an IAT (followed by a debrief) or more advanced training designs such as interactive workshops.

- UBT can be effective for reducing implicit bias, but it is unlikely to eliminate it.
- UBT interventions are not generally designed to reduce explicit bias and those that do aim to do so have yielded mixed results.
- Using the IAT and educating participants on unconscious bias theory is likely to increase awareness of and reduce implicit bias.
- The evidence for UBT's ability effectively to change behaviour is limited. Most of the evidence reviewed did not use valid measures of behaviour change.
- There is potential for back-firing effects when UBT participants are exposed to information that suggests stereotypes and biases are unchangeable.
- Evidence from the perspective of the subjects of bias, such as those with protected characteristics, is limited. This evidence could provide additional information on potential back-firing effects.

Awareness raising

The assessment indicates that awareness raising is the most likely aim of UBT and the aim most often achieved. Unconscious biases can be measured by a test such as the IAT. The evidence suggests that increasing the sophistication of UBT (for example by delivering an interactive workshop) can increase both participant awareness of their own implicit bias and concern about wider discrimination, and this awareness will continue to increase over time.

Reducing implicit bias

There is evidence that UBT reduces implicit bias, however, these biases are unlikely to be completely eradicated. When measured after the UBT, participants' scores on IATs are reduced but do not fall to neutral. Evidence suggests that more sophisticated UBT, such as those that combine awareness of unconscious bias, concern about its effects and the use of tools to reduce bias, can reduce unconscious bias up to eight weeks post-intervention.

Reducing explicit bias

Explicit bias change (shifting the attitudes and beliefs we have about a person or group on a conscious level) is harder to achieve through UBT than implicit bias. The evidence indicates three primary reasons for this:

- People tend to believe that they do not hold explicit prejudiced attitudes.
- Training participants who do hold explicit prejudiced attitudes are unlikely to disclose this. This is referred to as 'social desirability bias'.
- Most UBT interventions do not appear to be designed to alter or challenge explicit bias levels.

Changing biased behaviour

Evidence of behaviour change as an outcome of UBT is limited. Behaviour change is difficult to operationalise and measure, and therefore the evidence is harder to gather. For example, self-report measures of behavioural intentions are often described in studies, even though they do not actually tell us whether behaviour did change as a result of UBT.

Recommendations for practice

The evidence reviewed suggests that organisations should undertake a range of approaches to maximise the effectiveness of their UBT interventions.

Think about both UBT content and context

The content of a UBT intervention can influence its success in meeting its aim/aims. We recommend that organisations:

- Use an IAT, followed by a debrief session, to increase awareness of unconscious bias and to measure any changes in implicit bias.
- Deliver training to groups of people who work closely together (for example teams).
- Educate participants about unconscious bias theory rather than just providing information about the impact of unconscious bias using statistics.
- Include bias reduction strategies (such as promoting counter-stereotypic exemplars to challenge implicit stereotype endorsement and its effects) and bias mitigation strategies (such as more rigorous use of structured interviews to minimise the impact of bias), so that participants feel empowered to do something about unconscious bias.

Evaluate to measure effectiveness

We recommend that organisations:

- Are clear on the aim/aims of their UBT and use before-and-after measures to assess changes in, for example, awareness raising or attitude change.
- Randomly assign matched participants to intervention and control groups to evaluate the effectiveness of the training, and deliver the training to control participants once effectiveness has been established.
- Always carry out an evaluation after a UBT intervention to establish whether it has been effective in meeting its intended aim/aims.

Valid measures should be used to assess the effectiveness of the training. For example, if UBT has been designed for behaviour change, the evaluation should measure actual changes in behaviour, as opposed to behavioural intentions. The need for consistent and valid metrics for all aims of UBT is discussed in the 'further research' section.

See UBT as part of a wider programme

Finally, organisations should be aware of the limitations of UBT (including potential back-firing effects) and challenge underlying assumptions that raising awareness of unconscious bias or achieving short-term changes in implicit bias in isolation can lead to long-term change at an organisation level. For organisational level change to happen, organisational structures, policies and procedures must be targeted directly, perhaps overhauled. If the aim of UBT is to have an impact on company practice and employee behaviour to foster inclusive cultures where everyone meets their potential regardless of their identities (PwC, 2016; Nelson, 2017), UBT should be treated as just one part of a comprehensive strategy for achieving organisation-wide change.

Policy implications

Overall, this assessment is intended to promote better informed and evidence-based approaches to reducing inequalities in organisations, by interrogating the effectiveness of UBT. The findings raise the following questions about what role policy makers, government and employers should play in response to the findings:

1. Should a standard syllabus be compiled for the content of UBT?
2. Should the UK Governments and UBT experts work to develop a standardised set of methods for delivering UBT?

3. Should the UK Governments and UBT experts develop a nationally agreed metric or outcome measure to evaluate the effectiveness of diversity and inclusion initiatives, including UBT?
4. Should a 'What Works' network be created for the equality, diversity and inclusion agenda, within which UBT will comprise one strand?

Further research

The assessment found that only 18 sources of evidence were both relevant to the research question and adopted the minimum standards for quality research. The number of rigorous studies assessing the effectiveness of UBT is small and this is a significant finding in itself. More UK-based research and evaluations are required to strengthen the evidence base; further research should:

- systematically compare the impact of context (for example, organisations' strategic approaches to diversity), design characteristics (that is, the training content and delivery methods used), and effectiveness in reducing bias towards specific groups, and
- ensure valid measures of UBT aims (for example, adopt measures that assess actual behaviour change as opposed to asking participants about their intentions to change using questionnaires), with the aim of developing consistent and valid metrics for all aims of UBT, including awareness raising, implicit bias change, explicit bias change and behaviour change.

1 | Introduction

1.1 What is unconscious bias training?

Unconscious (or implicit) biases, unlike conscious biases, are the views and opinions that we are unaware of (Cornish and Jones, 2013); they are automatically activated and frequently operate outside conscious awareness (Lai *et al*., 2013) and affect our everyday behaviour and decision making (Kahneman, 2011). Our unconscious biases are influenced by our background, culture, context and personal experiences.

Workplace 'Unconscious Bias Training' (UBT) is a term used to describe a session, programme or intervention in which participants learn about unconscious bias, typically with a view to reducing the negative impact of bias on organisational practice and individual behaviour. UBT generally, although not exclusively, teaches employees about the negative impact of biases on people with protected characteristics, such as women or ethnic minorities. It is widely accepted that making people aware of their (unconscious) biases is the first step towards addressing the manifestation of them (Lee, 2017; Devine *et al*., 2012).

Unconscious bias

UBT is often designed, developed and modified on the basis of the large body of research on unconscious bias. During everyday interactions, our brains receive an influx of information. Unconscious biases arise because we rely on 'short-cuts' to filter this information rapidly. The function of these short-cuts, or heuristics, is to categorise and make decisions about people and tasks efficiently.

One of the negative consequences of this automatic processing is the influence of social stereotypes on our decision making. There is ample research documenting the influence of stereotypes on workplace evaluations and decision making (for example Eagly and Karau, 2002; Correll, 2017; Kossek *et al*., 2017), leading to detrimental outcomes for women, ethnic minorities, disabled people and others with a protected characteristic.

The aim of training

Given the negative impact of stereotypes on outcomes relating to equality, diversity and inclusion in the workplace, UBT is designed to increase awareness of unconscious bias through instruction. Participants learn that even where they do not consciously endorse a stereotype; it can influence their attitudes and behaviour outside their awareness. In addition to awareness raising, the aim of training is to teach methods to alleviate unconscious bias (Girod *et al*., 2016).

Organisations' ultimate purpose in implementing unconscious bias interventions such as training is to reduce or remove objective workplace inequalities (Nelson, 2017). However, the belief that UBT can achieve this rests on the underlying assumptions that: (1) such inequalities are caused, at least in part, by the behaviour of everyone in the organisation (rather than just those responsible for managing the organisation); that (2), changing or addressing biases and attitudes will change behaviour; and that (3) changing individual behaviour will change organisational outcomes. Furthermore, this causal chain assumes that organisations expect individuals will change their behaviour and that these variations will lead to organisational transformation even when organisational culture or climate do not reinforce the altered behaviour or may even counter the effect. Thus, when evaluating the evidence about UBT, it is important to bear in mind the causal assumptions underlying its popular adoption by organisations. It is also important from an evidence-based practice perspective to analyse evidence for the problem first and only then consider evidence for possible solutions (Atewologun, Cornish & Briner, 2017).

Overall, our search for evidence reveals that there are very few evaluations of UBT. This makes it difficult to assess the efficacy of current approaches. Furthermore, little is known about how UBT design and delivery affects its effectiveness. For example, although theoretical frameworks point to best practice UBT design (for example Moss-Racusin *et al*., 2014) to maximise organisational change (for example Nelson, 2017), an evaluation of these design features has not yet been conducted.

1.2 An assessment of UBT and its effectiveness

This assessment was commissioned by the Commission in 2017 to shed light on the evidence available for the effectiveness of UBT.

Increasing use of UBT in UK workplaces

In her 2017 review, 'Race in the Workplace', Baroness McGregor-Smith highlighted the 'structural, historical bias' that prevents ethnic minorities, women, disabled people and others from progressing in their careers. She recommended that the UK Government create a free, online unconscious bias training resource to tackle the unconscious bias that she describes as 'much more pervasive and more insidious than the overt racism that we associate with the 1970s' (McGregor-Smith, 2017, p.2).

In its official response to Baroness McGregor-Smith, the UK Government highlighted its increased use of UBT in the Civil Service (Business, Energy and Industrial Strategy, 2017). Private sector organisations have also implemented UBT in increasing numbers. Case studies analysed for this assessment indicate that tens of thousands of leaders and staff have undertaken the training nationally and internationally through in-house and online UBT programmes.

The rising number of organisations adopting unconscious bias training warrants an investigation into its usefulness. This is particularly necessary given that numerous reports have questioned the usefulness of UBT (Noon, 2018; Moss-Racusin *et al*., 2014; Kalev *et al*., 2006). Research has also suggested that UBT can even activate negative stereotypes or elicit negative reactions (Rudman and Glick, 2001; Dobbin *et al*., 2015).

Aims of the assessment

This assessment aims, first, to bring to light the evidence for the effectiveness of UBT, specifically meeting its declared aims. These aims are: awareness raising; implicit bias change; explicit bias change; and behaviour change.

Second, the assessment analyses the evidence to identify the boundary conditions within which UBT is deemed effective. Relatedly, evidence of when UBT may be less effective or, in fact have a negative impact, is required. Finally, this assessment aims to highlight the evidence gaps that will provide a pathway for further research.

1.3 Scope of the assessment

This assessment aims to inform the Commission's response to recommendations to implement UBT across the UK.

To meet the aims of this assessment, the broad research question is 'What is the evidence for the effectiveness of unconscious bias training (UBT)?'

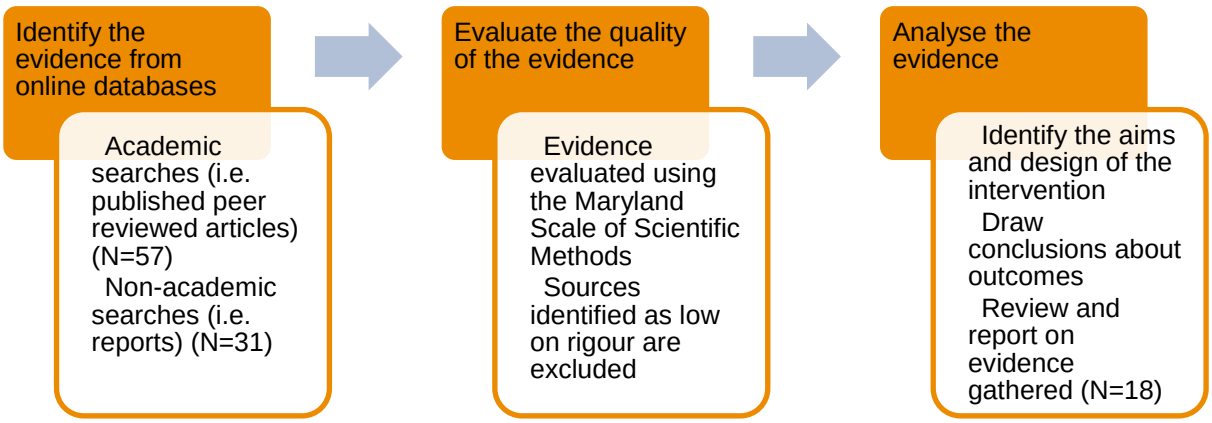
What is 'effectiveness'?

We refer to the 'effectiveness' of UBT in meeting its stated aims. Therefore, the effectiveness of the training interventions evaluated is dependent on the aims as set out (explicitly or inferred) by the training designer. A training intervention is deemed effective if robust evidence is provided that the intended aim/aims have been met as a result of the training. The assessment seeks to describe what works with regard to UBT, by summarising and analysing the evidence for what, how and for whom UBT is effective.

1.4 Methodology

A rapid evidence assessment was conducted (Figure 1.1). The three-stage process of collecting, evaluating and analysing the evidence is summarised here and detailed further in Appendix 1.

Figure 1.1 Methodology for rapid evidence assessment



To assess the methodological quality of the evidence, we used the Maryland Scale of Scientific Methods (MSSM) as recommended by the Civil Service in the Rapid Evidence Assessment Toolkit (Civil Service, 2014). For inclusion in this assessment, we selected studies that reached a minimum level of 2 according to the MSSM (Table 1.1; see Appendix 1 for further information).

Table 1.1 Assessment sources by level of rigour

Maryland Scale of Scientific Methods Level	Number of sources
2 – slightly rigorous	7
3 – moderately rigorous	1
4 – very rigorous	6
5 – extremely rigorous	4

2 | Key findings

2.1 Can unconscious bias training meet its aims?

As described in Chapter 1, this rapid evidence assessment sets out to evaluate the evidence for the effectiveness of unconscious bias training (UBT), where 'effectiveness' is determined as the extent to which UBT meets one or more of the four stated aims of training, as identified from the studies reviewed. This chapter considers the evidence that UBT is effective for a) awareness raising, b) implicit bias change, c) explicit bias change and d) behaviour change. In Boxes 1 to 4, we describe studies that are exemplars of each area of UBT aims.

Raising awareness of bias

Eleven of the studies used in this assessment explicitly stated that the aim of the UBT intervention was to raise awareness of unconscious bias. Hausmann et al. (2014) and Capers *et al*. (2017) showed self-reported increased awareness of unconscious bias after a UBT intervention. Further, Moss-Racusin et al. (2016) found an increase in participants' knowledge about diversity issues after their participation in a UBT workshop. Notably, this study did not measure bias awareness through mere self-report but through participants' heightened ability to detect accurately the gender diversity of their environment. We conclude that UBT interventions can increase awareness of bias.

Box 1. Awareness raising

(This study was rated at Maryland Scale of Scientific Methods (MSSM) Level 2, 'slightly rigorous')

In one US study, Whatley (2018) used an Implicit Association Test (IAT) and implemented a 'bias literacy workshop' (Carnes *et al*., 2012) to assess the effectiveness of UBT on multi-disciplinary staff team members for awareness raising and behaviour change towards African American students in special education. The bias literacy workshop involved educating participants on the theory of unconscious bias and teaching strategies to mitigate bias. Using a pre- and post-test design,

Whatley measured participants' scores on 'objectivity' and 'teacher expectations'. Results showed significant differences on both measures, indicating that staff increased awareness of their vulnerability to bias (objectivity measure) and had more positive expectations of individual students following the UBT intervention (teacher expectations measure).

Changing implicit bias

Eleven of the sources used in this assessment explicitly stated that the aim of the UBT intervention was to change implicit bias. Leslie *et al*.'s (2017) study found a reduction in implicit bias for second year medical students post-intervention (though this did not happen for first year students), and reductions for sexual orientation preference and racial preference (but not weight preference). Carnes *et al*. (2015) found no significant change in IAT scores after their unconscious bias workshop. However, Girod *et al*. (2016) found a reduction in implicit gender bias across all groups, including for men and older participants who held stronger biases. Further, Devine *et al*. (2012) found reduced race-IAT scores eight weeks after a habit-breaking intervention. This intervention is based on the premise that unconscious bias is like a habit that can be reduced through a combination of awareness of unconscious bias, concern about its effects, and the use of tools to reduce bias.

We conclude that there is mixed evidence for the effectiveness of UBT for reducing implicit bias. The results reported by Leslie *et al*. and Girod *et al*. suggest that UBT interventions can reduce the strength of the bias; however, we found no evidence to show that UBT can reduce bias to the extent that there is 'neutral' preference.

Box 2. Awareness raising and implicit bias change

(This study was rated at MSSM Level 2, 'slightly rigorous')

Girod *et al*. (2016) evaluated the impact of a 20-minute educational presentation on reducing gender bias in 281 faculty members from 13 clinical departments at Stanford University in the US. The study used pre- and post-test measures. Pre-intervention implicit measures (using the IAT) showed a slight preference for men in leadership positions (this bias was stronger in male, compared with female, participants, and older, compared with younger, participants). All racial groups showed the same extent of implicit gender bias pre-intervention. The implicit preference for men in leadership reduced immediately post-intervention, at statistically significant levels across gender, age and race. However, male and older participants still had stronger implicit biases.

Changing explicit bias

Nine of the sources used in this assessment measured explicit biased attitudes. Moss-Racusin *et al*. (2016) found a reduction in 'subtle gender bias' post-intervention. Participants responded to statements such as 'Discrimination against women is no longer a problem in the United States.' This measure could be considered a measure of explicit bias due to the self-report nature. However, Moss-Racusin *et al*. (2016) argue that these attitudes tend to exist beyond conscious awareness, that is, they are implicit as opposed to explicit biases.

After a UBT intervention, Carnes *et al*. (2015) showed an increase in self-reported attitudes that behaviours consistent with gender equity (for example promoting gender equality in their faculty department) will yield positive outcomes. However, Jackson *et al*. (2014) found that endorsement of explicit gender stereotypes was not reduced after a UBT intervention for men who held these explicit attitudes pre-intervention. Women did not hold explicit gender stereotypes about women in science, technology, engineering and mathematics (STEM) pre-intervention.

Overall, the evidence for UBT's effectiveness for changing explicit bias is weaker than that of either awareness raising or of changing implicit bias. Further, it is unclear from the available research how best to measure explicit bias and how to interpret findings vis-a-vis implicit bias changes. A recent meta-analysis suggests that IAT results have low correlations with explicit measures of bias for several measures of discrimination (Oswald et al., 2013). We discuss some of the challenges of assessing explicit bias using UB-related interventions in Chapter 5.

Box 3. Awareness raising and explicit bias change

(This study was rated at MSSM Level 2, 'slightly rigorous')

Moss-Racusin *et al*. (2016) implemented a two-hour 'scientific diversity' workshop for 126 life sciences instructors at sessions across the US. The workshop was designed following Moss-Racusin *et al*.'s (2014) framework identifying four design elements for effective design interventions (for a review, see Moss-Racusin *et al*., 2014). The aim of the workshop was to a) increase awareness of gender diversity issues, b) reduce gender bias and c) increase action readiness for diversity-related issues. Self-report evaluation showed that, post-intervention, participants were more aware of diversity issues, indicated less gender bias towards women in STEM and increased readiness to take action on diversity issues.

Changing behaviour

Ten of the sources used in this assessment stated that at least one aim of their UBT intervention was behaviour change. However, only two of the studies measured actual behaviour change.

Moss-Racusin *et al*. (2016) measured 'promotion-focus' and 'prevention-focus' as indicators of 'action readiness' via self-report statements including 'Right at this minute, in terms of my approach to diversity, I'm feeling… free to pursue my goals/ confident that I can go after my goals/ focused on what I will achieve.' Although research has shown that the use of promotion-focused indicators is associated with performance outcomes, there is insufficient evidence that these statements indicate behaviour change resulting from participating in UBT activities.

Sweetman's (2017) evaluation of a UBT workshop with higher education staff recorded an increase in participants' self-reported pro-equality motivation and 'action tendencies'. Again, this falls short of providing evidence of actual behaviour change.

Carnes *et al*. (2015) found no difference in action but did find a relationship between the proportion of a department attending the UBT workshop and departmental action at three months. This is an indication of group-level, rather than individual-level behaviour change. By training people in teams, the team is able to work together to implement changes to group activities and members can support each other to embed these changes; this is likely to increase the impact of the training, compared with only individual-level behaviour change.

Capers *et al*. (2017) measured the effects of taking an IAT on 140 members of a medical school's admissions team's a) awareness about their scores and b) subsequent admissions decisions regarding race. On average, there was an implicit White preference in the IAT prior to the admissions process. In the follow-up survey post-admissions, 67 per cent reported that the IAT might be helpful in reducing bias and 48 per cent reported that they were conscious of their results in the interview process. Further, 21 per cent reported that knowing their IAT result affected their admissions decisions. These responses indicate awareness raising. Capers *et al*. (2017) reported that the cohort of students following the admissions team's participation in the IAT was the most diverse in the medical school's history. Although the researchers attributed this increase to the admissions team's behaviour change towards the applicants, the statistics revealed that no more places were offered by staff to ethnic minorities following the intervention; rather, more ethnic minority students chose to accept their places. The researchers suggested that this outcome was likely due to participants' behaviour change in a positive way towards

the prospective students; however, this explanation is not supported by any further research from this particular study.

We conclude that there is insufficient evidence to indicate that UBT is effective for behaviour change. This is for two reasons: 1) the research examining behaviour change is limited, and 2) methods for evaluating behaviour change mostly have low validity, in that they do not measure actual observed change. An example of behaviour change is outlined below (Box 4).

Box 4. Behaviour change

(This study was rated at MSSM Level 4, 'very rigorous')

Forscher *et al*. (2017), in a US study, implemented a two-month 'habit-breaking UBT intervention with 292 students at a US university based on Devine *et al*. (2012). This intervention is based on the premise that unconscious bias is like a habit that can be reduced through a combination of awareness of unconscious bias, concern about its effects, and the use of tools to reduce bias. The intervention involved an IAT, a test debrief and a combination of educational and training sessions on the theory of unconscious bias and how to mitigate its effects. Contrary to Devine *et al*.'s (2012) earlier findings, IAT scores reduced post-test for intervention and control participants, and the effects of awareness of unconscious bias reduced in the second-week post-intervention. However, a follow-up study with both the intervention and control group also indicated long-term behaviour change. Participants in the original intervention group were more likely to comment on a public forum objecting to an essay endorsing racial stereotyping compared with original control group participants. Although the researchers could not control for contamination during the two years following the intervention, this is the first study to examine such long-term effects of UBT and possible effects on actual behaviour.

2.2 Back-firing effects of UBT

Two of the studies that were used to inform this assessment indicated potential negative effects of UBT, although the design of both studies makes it difficult to draw firm conclusions. Hausmann *et al*.'s (2014) hospital-based study (see Box 5), found that patients treated by the UBT intervention group reported a decrease in positive ratings for their interactions with office staff compared with the control group (who, counterintuitively, yielded more positive patient responses). This finding could be

explained by the intervention group's post-intervention beliefs (gained from the training content) that stereotypes can be useful in a healthcare setting, and that biases cannot be changed. In the researchers' discussion of these results, they stress the need to: 'Emphasize in the training that unconscious bias can be changed and its impact on behaviour reduced, and introduce specific techniques for accomplishing this' (Hausmann *et al*., 2014, p.4).

The second study is Leslie *et al*. (2017) (described more fully in Chapter 3). Although the study reports a significant decrease in IAT scores pre- to post-intervention, indicating lowered implicit bias, this was not the case for all test groups. One group's average IAT scores did not reduce pre- to post-intervention. This finding might have been due to a low pre-test score in this group. It could, however, also be because the intervention was ineffective or actually had a negative impact on implicit biases. The design of the study does not allow for more specific conclusions to be drawn.

Box 5. Potential back-firing effects

(This study was rated at MSSM Level 4, 'very rigorous')

Hausmann *et al*. (2014) implemented a UBT programme for patient aligned care teams in the US to reduce the negative effect of unconscious bias on veteran patient care. Participants were 18 randomly allocated teams. Nine were placed in the programme condition and nine were comparisons. Training was delivered in three parts using online training, workshops and practical exercises. After 30 days, participants self-reported that the training was helpful (57 per cent), and it had made them aware of how their biases affect patient care (68 per cent). Almost half reported that they had a) created new habits to explore unconscious biases (45 per cent), b) noticed a positive change in the way that their patients responded to practice (41 per cent), and, c) shared what they had learnt with co-workers and trainees (41 per cent). Agreement with these statements was similar after 90 days. A total of 91 per cent and 88 per cent of participants indicated that they believed they had been successful in applying the training in their clinical or administrative practice at the 30-day and 90-day time-points, respectively. Patient evaluations showed a slight positive increase of their experience of clinical staff in respect to communication, comprehensiveness, and self-management support but this was the case for the control group as well. Patients reported a decrease in positive ratings for the interaction with office staff compared with the control group (who yielded more positive responses). Intervention participants were more likely to agree that stereotypes can be helpful and can be used in a healthcare setting, and that biases cannot be changed. The researchers suggest this was likely due to participants learning about the ubiquity of cognitive processes that lead to stereotypes and biases.

2.3 Summary

Overall, our evaluation of rigorous studies on the effectiveness of UBT indicates the following:

- UBT is effective for awareness raising by using an IAT (followed by a debrief) or more advanced training designs such as interactive workshops.
- UBT can be effective for reducing implicit bias, but it is unlikely to eliminate it completely.

- UBT interventions are not generally designed to reduce explicit bias and those that aim to do so have yielded mixed results.
- Using the IAT and educating participants on unconscious bias theory is likely to increase awareness of and reduce implicit bias.
- The evidence for UBT's ability effectively to change behaviour is limited. Most of the evidence available does not adopt valid measures of behaviour change.
- Only one experimental study examined long-term behaviour change and this UBT provided no control over participants in the two-year period between the intervention and behavioural measure.
- There is a positive relationship between higher department attendance rates for UBT, mandatory UBT and UBT as part of a broader organisational diversity strategy and behaviour change, although this evidence is not rigorous. Design considerations for UBT are discussed further in Chapter 3.
- There is potential for back-firing effects when participants are exposed to information that suggests stereotypes and biases are unchangeable.
- Evidence from the perspective of the subjects of bias, such as those with protected characteristics, is limited. This evidence could provide additional information on potential back-firing effects.

3 | Design considerations of UBT

By ‘design considerations’, we mean the contextual factors and design characteristics of UBT that are often considered influential components of the effectiveness of the training (Nelson, 2017; Bezrukova *et al*., 2016). This chapter looks in detail at whether design considerations for delivering UBT are effective in the four aims of a) awareness raising, b) implicit bias change, c) explicit bias change and d) behaviour change.

3.1 Contextual factors

Bezrukova et al. (2016) suggest that research has overlooked the role of context in effective diversity training programmes. ‘Contextual factors’ refer to the organisational climate/environment and approach to implementing UBT. Two examples would be implementing mandatory training as opposed to voluntary training; and integrating UBT alongside other diversity initiatives, as opposed to a standalone UBT intervention. In this section, we explore the evidence available to suggest that the contextual factors, or overall approach taken when implementing UBT in an organisation, may have an impact on its effectiveness in awareness raising, implicit and explicit bias change and behaviour change.

Mandatory versus voluntary training

We did not find any rigorous studies on the impact of mandatory versus voluntary attendance on the outcomes of UBT. One large-scale UK-based study (Business in the Community, 2012; 2013) found positive relationships between mandatory training for recruitment staff and hiring outcomes for women and ethnic minorities. We highlight this study, given its large data set. However, the correlational nature of the findings limits the degree of confidence we can have about causality.

There is debate in the wider research on diversity training about the benefits of mandatory versus voluntary training (for example Kalev *et al*., 2006; Bezrukova *et al*., 2016). Bezrukova and colleagues suggest these two approaches yield different

outcomes – voluntary training is more likely to be positively received by employees, whereas mandatory training is more likely to lead to behaviour change. However, Bezrukova *et al*.'s (2016) focus on general 'diversity training' is not specific enough to UBT to draw conclusions about the impact of mandatory and voluntary UBT on any outcomes.

Box 6. Mandatory UBT for behaviour change

(This study was rated at MSSM Level 1, 'low rigour')

The 2013 Business in the Community Benchmark Trend included 98 UK organisations with a combined workforce of almost two million people. The results from this benchmark report showed a positive correlation between mandatory diversity training and successful hiring of women. Business in the Community's 2012 benchmark report showed that 50 per cent of organisations with higher recruitment of ethnic minorities implemented mandatory UBT for recruitment staff, compared with only 5 per cent of organisations with low recruitment of ethnic minorities. Furthermore, organisations in which women were as successful at being hired as they were shortlisted were eight times more likely to have mandatory UBT for recruitment staff.

Participant demographics and work roles

Jackson *et al*.'s (2014) gender diversity training to reduce gender bias in science, technology, engineering and mathematics (STEM) found that personal implicit associations were reduced after the training for men but not for women. A personal implicit association requires participants to group categories such as 'female scientist/engineer' with attributes such as 'I like' or 'I don't like.' In contrast, a standard Implicit Association Test would not be personalised and ask respondents to rate using attributes such as 'pleasant' or 'unpleasant'.

Moss-Racusin *et al*. (2016) found that men held stronger subtle gender biases than women at the pre-test stage of their 'scientific diversity' workshop, a gap that was reduced to a non-significant difference between men and women at the post-test stage. Girod *et al*. (2016) found stronger implicit biases for men and older people prior to their educational intervention. Although post-intervention men and older people's implicit biases were stronger than those of women and younger people, the intervention yielded similar decreases across the groups. Jackson *et al*. (2014) and Moss-Racusin *et al*. (2016) both delivered UBT via interactive workshops, whereas Girod *et al*. (2016) delivered UBT as a lecture. The difference in methods of delivery may explain the difference in findings.

The evidence for participant demographics suggests that implicit bias reduces for both male and female, and older and younger participants after a UBT intervention, even when men and older people hold stronger implicit biases pre-intervention.

Box 7. Impact of training with participants from the same team

(This study was rated at MSSM Level 5, 'extremely rigorous')

Carnes *et al*. (2015) implemented a two and a half hour unconscious bias interactive workshop to 92 departments at the University of Wisconsin in the US. The design used randomised sampling and pre-test and post-test comparisons. The aim of the workshop was to a) increase awareness of gender bias in academia and b) promote motivation, self-efficacy and positive outcome expectations for habitually acting in gender-equity consistent ways. The workshop educated participants about the 'business case' for diversity, and theory and research on unconscious bias. Three modules were run. Module one outlined the nature of bias as a 'habit'. Module two promoted 'bias literacy,' which described and explained the different forms of bias and how they manifest. Module three promoted self-efficacy, which provided evidence-based strategies for overcoming bias. At three days, the experimental group reported greater increases in personal bias awareness, motivation, self-efficacy and expectations of positive outcomes for behaving in gender-equity consistent ways. At three months, personal bias awareness and self-efficacy persisted. The experimental group showed increased external motivation. No change in IAT associating 'male' with 'leader' and 'female' with 'supporter' was found. No differences in action were found. However, the experimenters reported that when at least 25 per cent of a department faculty attended the workshop, there was a significant increase in action at three months but department leader attendance had no effect. This suggests that delivering UBT sessions to work teams may be more effective than open sessions for individual participation.

3.2 Design characteristics

'Design characteristics' refer to the features of the UB training itself, for example, the methods of delivery such as online or face-to-face training, as well as the content of the unconscious bias training, that is, what information participants receive as part of the training. In this section, we explore the evidence available to suggest that the design characteristics of UBT may have an impact on effectiveness in awareness raising, implicit and explicit bias change and behaviour change.

Method of delivery

Box 8. Comparing methods of delivery of UBT

(This study was rated at MSSM Level 3, 'moderately rigorous')

Google (2013) conducted an experiment in the US to evaluate whether their UBT workshop met the training's aims of increased awareness and understanding of unconscious bias and motivation to overcome it. Participants were randomly allocated to one of three groups: participation in a live workshop, online self-study video of the workshop, or no UBT (control group). A self-report survey was used to measure participants' awareness and understanding of unconscious bias and motivation to overcome it. Participants' awareness and understanding of unconscious bias and motivation to overcome it were significantly higher post-test compared with pre-test in both online and face-to-face groups, compared with the control group. Results persisted one month after the workshop. Notably, face-to-face training did not yield stronger effects than online training.

Box 9. Online UBT

(This study was rated at MSSM Level 2, 'slightly rigorous')

A UK case study example of online UBT is PwC's (2016) mandatory *Open Mind* e-learning tool. The UBT intervention is designed to make employees aware of the impact of their unconscious biases and to take action to be more open-minded. Evaluation of the training, based on participant self-report feedback, showed the training increased participants' awareness of their own biases. Further, PwC reported a significant increase in the diversity of their graduate hires in 2016: female and Black, Asian and Minority Ethnic (BAME) representation reached 43 per cent and 34 per cent of hires respectively (compared with 37 per cent and 25 per cent in 2012). PwC concluded that UBT is effective for driving employee behaviour change and an inclusive culture, fairer promotions and appraisals processes, which ultimately led to the appointment of two women to the executive board (Business in the Community, 2012). Without a control comparison group, however, it is difficult to attribute these changes to the method of implementation (or any of the other interventions PwC had in place to improve diversity). Despite this, the study is commendable in that it was based on employee data (rather than laboratory or student samples) and used before-and-after comparisons. Thus, this case study offers encouraging support for the effectiveness of online training for driving behaviour change.

Moss-Racusin *et al*. (2014) designed a framework based on the evidence available for prejudice reduction strategies. They suggest that training design should use active learning techniques such as writing and speaking. Active learning helps participants to engage with the course content (for example, through problem-solving and group discussion). Moss-Racusin *et al*. (2016) used this framework in their scientific diversity workshop and found support for their training aims (increased awareness, reduced gender bias and action readiness). However, without a systematic comparison of approaches and evidence of causality, we cannot conclude that interactive workshops are more effective than less interactive lecture-based learning. Furthermore, as Moss-Racusin and colleagues' (2014) framework covers the four areas of aims, it is not clear which aim or outcome (awareness raising, change in implicit or explicit bias or behaviour change) the interactive element targets.

Training content

Capers *et al*. (2017) found that using a single IAT increased participants' self-reported awareness about their biases and impact on their behaviour. Also, Leslie *et al*. (2017) found that medical students who took an IAT, followed by a test debrief, and underwent an educational curriculum on equality in healthcare, had significantly lower IAT scores six months post-intervention than a comparison group who took no pre-intervention IAT and only received the educational curriculum intervention. Also, informing participants about unconscious bias theory has been found to be a more effective content of training compared with informing them about the negative impact of UBT (Repelaer van Driel, 2015).

Box 10. Comparing content of UBT

(This study was rated at MSSM Level 2, 'slightly rigorous')

Leslie *et al*. (2017) evaluated the effectiveness of two different types of UBT content for reducing implicit bias: an IAT intervention (with a debrief) and a curriculum focused on equality in healthcare. The interventions targeted sexual orientation, race and weight biases among medical students at the Louisville School of Medicine in the US. The researchers used randomised semi-control trials and pre-test and post-test comparisons. One group who experienced both interventions showed a reduction in IAT scores between pre-test and post-test for sexual orientation and race. The preference for 'straight' and 'White' remained but reduced towards neutral. Another group who experienced both interventions showed no significant reduction in their

implicit preference for ‘straight,’ ‘White’ or ‘thin’. However, their results indicated lower IAT scores than a third group who undertook only the academic curriculum, for sexual orientation and racial bias (but not weight).

Box 11. Content theory versus impact of UBT

(This study was rated at MSSM Level 4, ‘very rigorous’)

Repelaer van Driel (2015) directly compared learning about unconscious bias theory with learning about the impact of unconscious bias in a randomised study of 176 students and employees at a Dutch university. The experiment used a mock hiring paradigm, asking participants to evaluate real assistant professor applications, imagining that they were an employer. Participants were randomly assigned to conditions where, prior to applicant evaluation, they read information on either: the under-representation of women in academia (the impact of unconscious bias via statistics), implicit gender bias (theory of unconscious bias), both, or neither (control). After the information, participants all read the same application that was either a male or female candidate. Participants who had received information about implicit gender bias rated the female candidate as more competent and hireable than those who had not. Where participants were given information on the impact of unconscious bias via statistics, women were rated only slightly more competent than those who did not receive an intervention (although this difference reached statistical significance). Receiving both forms of content (statistical and theoretical information) did not reduce bias any more than receiving information on theory did alone. Men were rated as less competent and hireable, shifting towards a bias against male candidates. This study indicates that educating participants about unconscious bias theory is more effective than using statistics to illustrate the impact of UBT for reducing gender bias in hiring.

Methods for mitigating unconscious bias

Methods for reducing the negative impact of unconscious bias (UB) fall into two categories: bias mitigation and bias reduction strategies. Bias mitigation strategies enable people to limit UB’s negative impact but may not change the actual level of bias (for example, techniques designed to make selection and promotion decisions more objective; Isaac *et al*., 2009). Bias reduction strategies are scientific, evidence-based methods for decreasing levels of implicit bias (for example, challenging participants’ unconscious negative thinking by presenting positive counter-stereotypic images). In their rapid evidence assessment Cornish and Jones (2013) identify a range of bias reduction strategies from the scientific research:

- Discounting commonly held stereotypes using positive and counter stereotypic images.
- Changing how an outgroup member is evaluated and categorised through the use of evaluative conditioning.
    - Here, participants are exposed to repeated pairing of images of outgroup members with positive images but in a way that disguises the purpose of the activity from participants.
- Increasing contact between different groups to change the level of threat evoked in the presence of an outgroup member.
- Encouraging people to take responsibility for their implicit biases by using cognitive strategies such as implementation intentions (if-then action plans) and appropriate attributions for outgroup behaviour.
- Encouraging participants to choose valuing diversity freely rather than through fear of external sanction, or choosing a multicultural, rather than a colour-blind, approach to diversity.

Lai et al. (2014) conducted a large-scale study using bias reduction strategies to reduce implicit racial bias. They compared 17 bias reduction strategies on a total of 17,021 US citizens recruited via the Project Implicit website. The researchers reported that the three most effective clusters of strategies for reducing implicit bias were: exposure to counter-stereotypic exemplars, intentional strategies to overcome bias, and evaluative conditioning. Lai et al. conducted a later (2016) study to gauge the relative effectiveness of different bias reduction strategies. Participants were 827 American university undergraduates. Participants (some of whom took a pre-test race IAT) were randomly assigned to one of nine conditions: eight bias reduction strategies and one control group. Similar to their 2014 findings, the researchers found no effect of intervention condition on explicit racial preferences. Also in line with the original findings, some bias reduction strategies were successful in reducing implicit racial bias. Five interventions yielded significantly lower IAT scores than the control group (indicating less implicit bias after conducting the bias mitigation strategy):

- experiencing a vivid counter-stereotypic scenario
- practicing an IAT with counter-stereotypic exemplars
- evaluative conditioning with the Go/NoGo-Association Task
- using implementation intentions, and
- ‘faking’ the IAT.

However, none of the intervention IAT scores were lower (that is, less biased) than the control group's scores at follow-up (on average three days), suggesting that implicit biases were reduced only temporarily. Furthermore, participants still had an overall implicit preference for 'White' over 'Black' at pre-test, post-test and follow-up. Thus, these bias reduction interventions appear ineffective for changing implicit racial bias over the long term. Neither do they affect explicit bias.

Kawakami *et al*. (2005; as cited by Cornish and Jones, 2013) found that matching female faces to non-stereotypic words reduced implicit gender bias. However, a follow-up hiring study found no difference between the experimental and control conditions in discrimination towards female candidates. Discrimination in hiring female candidates reduced only when the researchers used a filler task (used as a distraction, to disguise the true purpose of the activity) or a task designed to minimise conscious control. The researchers suggest that conscious awareness created a backlash. This is because when people believe that they have been influenced, they will moderate their response to oppose the perceived influence (Wegener and Petty, 1997; Wilson and Brekke, 1994). Thus, conscious awareness of bias mitigation strategies may backfire and undo the effects (unless participants are in agreement with the direction of the influence).

Box 12. Mindfulness as a method of bias reduction

(This study was rated at MSSM Level 5, 'extremely rigorous')

Lueke and Gibson (2015) used experimental methods to examine the effectiveness of mindfulness for reducing implicit bias. The experiment was conducted with 72 college students at Midwestern University in the US. The researchers used an experimental and control design. Half the participants were randomly allocated to the mindfulness condition in which they listened to an audio for 10 minutes that instructed them to become aware of their bodily sensations. Control condition participants listened to a 10-minute audio about natural history. Post-intervention IATs showed lower implicit race and age bias in the mindfulness group compared with the control group. Furthermore, there was a significantly lower Black/bad and old/bad association for participants in the mindfulness condition. These findings suggest that mindfulness may be an effective bias mitigation strategy for reducing negative implicit associations of race and age.

Box 13. Mindfulness as a method of bias reduction for behaviour change

(This study was rated at MSSM Level 5, 'extremely rigorous')

Lueke and Gibson (2016) replicated their 2015 mindfulness study (as described above). However, to increase the rigor of their methodology, they adopted an extra control 'attention' condition as well as a 'pure' control condition. This extra control condition was included to eradicate 'focused attention' as an explanation for reduced bias. This enabled the researchers to attribute findings to the content of the mindfulness audio. To avoid making racial stereotypes salient, an IAT was not used. Instead, participants completed a 'trust game' task. The game requires participants to allocate $0-$10 of their allocated $50 to either Black, White, Asian or Middle Eastern 'participants', for whom they see faces only. Participants 'trust' that the other 'participant,' has recorded equally generous allocations. Participants were told that the participant with the most hypothetical money at the end of the game will win $20. The researchers measured equal trust in Black and White 'participants' for the intervention condition but a White-preference in both control conditions. This suggests that mindfulness was effective for reducing implicit racial bias. Further, the fact that participants had the opportunity to win $20 suggests that mindfulness even affected behavioural decisions in the 'trust game'.

Training duration

Box 14. Longer training duration

(This study was rated at MSSM Level 4, 'very rigorous')

Devine *et al*. (2012) conducted a 'prejudice habit-breaking' intervention over 12 weeks with students in the US. The aim of the intervention was to increase implicit bias awareness, increase concern about the effects of that bias, and apply strategies to reduce implicit bias. Participants were randomly assigned to the intervention or control groups. All participants completed IATs and received IAT debriefs. Intervention participants received 45 minutes of interactive educational and training sections, including theory on unconscious bias, the impact of unconscious bias and strategies for reducing implicit racial bias. Implicit racial biases were reduced four weeks after the intervention and remained low eight weeks post-intervention. No explicit bias change was measured, but concern for the impact of implicit bias increased over time.

3.3 Summary

In summary, with regard to design considerations in implementing UBT, the evidence suggests that:

- Male participants hold stronger unconscious gender biases than female participants but this gap can be reduced using UBT, suggesting that UBT may be more effective for men, compared with women, in relation to specific gender biases.
- The evidence is mixed for the strength of the effects of UBT on reducing bias in different gender or age groups. Thus, design characteristics may differentially influence how various groups are affected by UBT.
- Mandatory UBT is more likely to be effective for behaviour change than voluntary UBT, although this is not supported by rigorous studies.
- There is some evidence that online and face-to-face UBT are equally effective for awareness raising.
- Longer UBT is more likely to increase awareness, reduce implicit and explicit bias and change behaviour compared with shorter training.
- The IAT (followed by a test debrief session) can be used to increase awareness about unconscious bias, and may be beneficial to reduce but not eliminate implicit bias.
- Bias reduction strategies are effective for reducing implicit bias but are ineffective for reducing explicit bias (although this is not unexpected given the nature of explicit bias).
- There is evidence that a mindfulness intervention can reduce implicit bias, and these effects may extend to reduce discriminatory decisions.
- Bias mitigation strategies may have a back-firing effect if participants are not in agreement with being influenced or with the direction of influence.

4| Applying UBT to protected characteristics

4.1 Evidence for protected characteristics

Specific protected characteristics, for example gender, race and sexual orientation, are often the subject of unconscious bias training (UBT). In this chapter, the evidence is re-examined to draw conclusions about the effectiveness of UBT for reducing bias towards members of particular protected groups.

Good quality evidence from corporate interventions is limited. Only one intervention directly compared outcomes for different protected groups (Leslie *et al*., 2017).

Gender

An Implicit Association Test (IAT) and short presentation reduced gender implicit bias in men and women, and older and younger people (Girod *et al*., 2016). Moss-Racusin *et al*.'s (2016) two-hour scientific diversity workshop increased awareness of gender bias, reduced self-reported gender bias and increased self-reported behavioural intentions to tackle issues of gender diversity. Although this study's gender bias measure did assess explicit bias, it is difficult to conclude that the workshop changed gender-biased behaviours, as self-report behavioural intentions are not evidence of behaviour change and can only indicate behavioural intention.

Carnes *et al*. (2015) found that behaviour change towards gender equality can occur at the department level, indicated by a relationship between the proportion of staff who attended the UBT and departmental action three months later.

Thus, there is evidence that training for gender bias can make participants more aware of their gender bias, reduce implicit bias and explicit bias. Furthermore, evidence indicates that UBT may lead to behaviour change towards gender equality. However, more evidence at the individual level that does not rely on self-report is required.

Race and ethnicity

Racial and ethnic minority bias is the second most frequent target of UBT (after gender). There is evidence that training for racial and ethnic minority bias can make participants more aware of their racial bias and reduce implicit racial bias.

A single race IAT increased a medical school's admissions team's awareness about their own biases (Capers *et al*., 2017). Some participants reported that the IAT had an impact on their decisions, and the school had the most diverse student cohort in its history following the IAT. The evidence for behaviour change is weak, but strong for raised racial bias awareness. Furthermore, Leslie *et al*. (2017) found a reduction in implicit racial bias after a lesbian, gay, bisexual and transgender (LGBT) and health-equity curriculum, suggesting that implicit racial bias can be reduced via UBT aimed at different protected groups. Lueke and Gibson (2015) and Devine *et al*. (2012) also demonstrated reduced implicit racial bias after UBT habit-breaking interventions.

However, evidence for the effects on explicit racial bias is limited. For example, Lai *et al*. (2014; 2016) found that explicit racial bias did not decrease following implicit bias reduction strategies.

Overall, there is encouraging evidence indicating that UBT may lead to behaviour change towards greater ethnic minority equality, but this needs to be examined further with rigorous methods in real-life settings.

Other protected groups

Leslie *et al*.'s (2017) LGBT and health-equity curriculum (a module teaching medical students about providing equal treatment to all patients) reduced implicit bias with regard to sexual orientation as well as race, but not weight. Only two interventions from this review addressed age bias, providing evidence that mindfulness could reduce implicit age bias (Lueke and Gibson, 2015; 2016).

Evidence of UBT's effectiveness for groups other than gender or race is limited to reductions in implicit bias. Further research is required to examine UBT's effectiveness in raising awareness, and changing explicit bias and behaviour for diversity dimensions beyond gender and race.

So far, research indicates only that implicit bias can be reduced for sexual orientation and age bias. There is no evidence for awareness raising, explicit bias change or behaviour change. More research is required to examine whether the results that have been found for gender and race can also apply to other protected characteristics.

4.2 Summary

In summary, with regard to the effectiveness of UBT on specific groups:

- Evidence shows that UBT can effectively raise awareness of gender bias and can reduce implicit and explicit gender bias.
- More data are required to evidence behaviour change towards gender equality, although the research suggests that action can occur at the group level.
- A single race IAT can raise participants' awareness about racial bias, and implicit racially biased attitudes are likely to be reduced through UBT.
- Evidence for explicit bias change is too limited but suggests that reducing explicit racial bias through UBT is difficult.
- Additional rigorous evaluations are required to conclude that behaviour change towards racial equality can occur, but there is encouraging evidence that this may be possible.
- There are a limited number of UBT interventions for other social identities. Therefore, more rigorous evaluations of UBT are required to assess its effectiveness for raising awareness and reducing bias towards other protected categories beyond gender and race.

5| Conclusion

This rapid evidence assessment was commissioned by the Commission to analyse and summarise the evidence for what, how and for whom unconscious bias training (UBT) is effective.

5.1 Main findings

Overall, our evaluation of rigorous studies on the effectiveness of UBT indicates a mixed picture and a need for further research to determine the effectiveness of unconscious bias training. We found that:

- UBT is effective for awareness raising by using an IAT (followed by a debrief) or more advanced training designs such as interactive workshops.
- UBT can be effective for reducing implicit bias, but it is unlikely to eliminate it.
- UBT interventions are not generally designed to reduce explicit bias and those that aim to do so have yielded mixed results.
- Using the IAT and educating participants on unconscious bias theory is likely to increase awareness of and reduce implicit bias.
- The evidence for UBT's ability effectively to change behaviour is limited. Most of the studies reviewed did not use valid measures of behaviour change.
- There is potential for back-firing effects when UBT participants are exposed to information that suggests stereotypes and biases are unchangeable.
- Evidence from the perspective of the subjects of bias, such as those with protected characteristics, is limited. This evidence could provide additional information on potential back-firing effects.

Awareness raising

The evidence suggests that raising awareness of bias is the most likely aim of UBT and the aim most often achieved. This can occur from taking a single IAT, through to participating in interactive workshops designed to reduce prejudice. All eleven of the awareness-raising interventions evaluated for this assessment (thus featuring

rigorous methodological designs) met their aims. Evidence further indicates that awareness raising is possible across delivery methods, that is, online and face-to-face, and effects can be long-lasting. However, all UBT evaluations should measure and manage bias awareness and, ideally, measure the long-term effects of awareness raising on bias. This is because teaching participants about stereotypes has the potential for back-firing effects by making negative stereotypes more accessible. Further study of this effect is needed.

Implicit bias

The evidence for UBT changing short-term implicit bias is consistent. Implicit bias is likely to be reduced, but not eradicated, through UBT. Bias reduction strategies such as using counter-stereotypic exemplars and evaluative conditioning effectively reduce implicit bias. Mindfulness has also shown to be an effective strategy for reducing implicit bias and potentially discriminatory decision making. However, these strategies are not commonly used in short-term UBT and the observed effects so far are mostly restricted to academic lab experiments. However, where they have been used in long-term habit-breaking type interventions (Devine *et al*., 2012), they have had positive effects for implicit bias reduction. Implicit bias reduction can have effects across all participant groups but appears to be more effective for those with stronger implicit biases pre-intervention.

Of the 11 studies which aimed to reduce implicit bias, eight were rated level four or five on the MSSM. There is strong evidence that implicit biases can be reduced eight weeks post-intervention when a sophisticated, habit-breaking design that is long-term and includes awareness-raising and bias mitigation strategies is used. Given the link between implicit bias and prejudiced behaviour (Rooth, 2010), the findings for sophisticated UBT designs for reducing implicit biases in the long term are promising.

Explicit bias

Compared with implicit bias, the evidence suggests that UBT is less likely to be effective in changing explicit bias. This evidence is mostly based on measuring before-and-after changes in self-reported attitudes. As expected, due to the nature of explicit bias, change in explicit bias is unlikely to occur as a result of UBT.

There are three main reasons why UBT interventions are unlikely to measure a change in explicit bias.

- People tend to believe that they do not hold explicit prejudiced attitudes.

- Training participants who do hold explicit prejudiced attitudes are unlikely to disclose this. This is referred to as 'social desirability bias'.
- Most UBT interventions do not appear to be designed to alter or challenge explicit bias levels.
    - The content of UBT typically focuses on informing participants of how implicit biases are activated in interpersonal and social interactions and that these will often be counter to their beliefs about their explicit bias.
    - However, many interventions do include measures of explicit bias to determine whether the intervention has had any effect on levels of explicit bias.

Behaviour change

The rigorous studies selected for this assessment do not in the main use highly valid measures of behaviour change. Behaviour change is often measured using self-report assessments of behavioural intentions for real-life interventions (that is, people saying they will act rather than independent observations of them acting). For example, hiring decisions, involvement in diversity-related action or interpersonal behaviour can all indicate behaviour change post-UBT but are sparse in the literature. Thus, this evidence of behaviour change as a result of UBT is weak and limited. We acknowledge, however, the difficulty in operationalising and measuring actual behaviour change.

Recent evidence does suggest that sophisticated habit-breaking interventions can have a long-term impact on behaviour, up to two years post-intervention (Forscher *et al*., 2017). Further studies incorporating frequent assessments, using relevant measures over extended time periods (to chart changes), with control groups, would bolster these findings. Another method for measuring behaviour change in UBT participants is to measure how the subjects of bias perceive participants' behaviour change. Where this has been done, results have highlighted discrepancies between participants' beliefs about their behaviour and the beliefs of others (Hausmann *et al*., 2014).

5.2 Further research

The assessment has identified a number of gaps in the evidence on UBT. To enable us to draw more confident conclusions regarding the effectiveness of implementing UBT in workplaces, further research is required in the following areas:

More UK-based evidence

Only one UK-based academic source met our criteria for inclusion in this assessment. This indicates that there is little academic research conducted on UBT in the UK. North American (primarily US) research can provide an indication of likely effects, but the nature of racial bias in particular does not necessarily transfer to the UK context. We recommend that this review is read with this limitation in mind.

Systematic comparisons

The evidence base would benefit from systematic comparisons of methodology, content and impact on protected groups. The lack of research comparing approaches, design characteristics and outcomes for different groups makes it difficult to draw conclusions for best practice.

Rigorous evaluation

It is evident from our searches that UK public and private sector organisations are increasingly using, or intending to use, UBT. Although attempts to record its effectiveness are laudable, there is a lack of systematic and rigorous evaluation of UBT in the workplace. Organisations attending to more rigorous evaluation could fill our knowledge gap of whether and how UBT changes bias and behaviours in UK workplaces, where academic publications are currently lacking.

Valid measurement

We found that aims are generally measured in the same way, but that the validity of these measures is not necessarily confirmed. Awareness is measured using self-report; implicit bias is measured using an IAT; explicit and behaviour are measured using self-report. There would be greater confidence in the effectiveness of UBT if each aim were measured using multiple approaches. For example, awareness could be measured by asking about a diversity-related situation to gauge awareness as opposed to asking participants about their awareness directly (see Moss-Racusin *et al*., 2016). More innovative approaches to the IAT could be used such as Jackson *et al*.'s (2014) personalised IAT or the 'trust game' (Lueke and Gibson, 2016). Equally, measures of behaviour change are required to measure actual behaviour differences as opposed to intentions or motivation. Currently, real-life behaviour change is not well-examined as an outcome of UBT. However, a recent long-term UBT intervention has yielded promising results (Forscher *et al*., 2017). Perceptions of UBT participants' behaviour change from the perspective of others, such as the subjects of bias, may also provide evidence of behaviour change (Hausmann *et al*., 2014).

Development of valid 360 degree feedback type approaches would help to address concerns about the subjectivity of these measures.

5.4 Recommendations

Our recommendations for practice and research arising from this assessment are below.

Recommendations for practice

The evidence assessed suggests that to improve their UBT interventions to maximise effectiveness, organisations should undertake a range of approaches.

Think about the UBT content

What is included in a UBT intervention can influence its success in meeting its aim/aims. We recommend that organisations should:

- Use an IAT, followed by a debrief session, to increase awareness of unconscious bias and to measure any changes in implicit bias.
- Educate participants about unconscious bias theory rather than just providing information about the impact of unconscious bias using statistics.
- Include bias reduction strategies (such as promoting counter-stereotypic exemplars to challenge implicit stereotype endorsement and its effects) and bias mitigation strategies (such as more rigorous use of structured interviews to minimise the impact of bias) so that participants have increased confidence about their ability to manage unconscious bias following the training.

Think about the UBT context

Research on the conditions that support the transfer of training to the workplace identifies the importance of the work environment in supporting participants to embed their new learning (Burke and Hutchins, 2007). In the current review, training people in groups consisting of at least 25 per cent of their team mates improved the effectiveness of the training. Organisations are therefore encouraged to:

- Deliver training to groups of people who work closely together (for example teams).

Evaluate to measure effectiveness

We recommend that organisations:

- Are clear on the aim of their UBT and use before-and-after measures to assess changes in, for example, awareness raising and/or bias change.
- Randomly assign matched participants to intervention and control groups to evaluate the effectiveness of the training, and deliver the training to control participants once effectiveness has been established.
- Always carry out an evaluation after a UBT intervention to establish whether it has been effective in meeting its intended aim/aims.

Valid measures should be used to assess the effectiveness of the training. For example, if UBT has been designed for behaviour change, the evaluation should measure actual changes in behaviour, as opposed to behavioural intentions. The need for consistent and valid metrics for all aims of UBT is discussed in the 'further research' section.

See UBT as part of a wider programme

In addition to considering the content, context and evaluation of UBT, organisations should be aware of the limitations of training (all training not just UBT) to bring about organisational change. Diversity practitioners and champions need to challenge underlying assumptions that raising awareness of UB or achieving short-term changes in implicit bias in isolation can lead to long-term, organisation-wide change. For organisational level change to happen, structures, policies and procedures must be targeted directly, perhaps overhauled. If the aim of UBT is to have an impact on company practice and employee behaviour to foster inclusive cultures where everyone meets their potential regardless of their identities (PwC, 2016; Nelson, 2017), UBT should be treated as just one part of a comprehensive strategy for achieving organisation-wide change. UBT should be treated as one step towards achieving organisational change, through awareness raising, implicit bias change and motivation to act.

Policy implications

Overall, this assessment is intended to promote better informed and evidence-based approaches to reducing inequalities in organisations, by interrogating the effectiveness of UBT. The findings also raise the following questions about what role policy makers, government and employers should play in response to the findings.

1. Should a standard syllabus be compiled for the content of UBT?

2. Should the UK Government and UBT experts work to develop a standardised set of methods for delivering UBT?
3. Should the UK Government and UBT experts develop a nationally agreed metric or outcome measure to evaluate the effectiveness of diversity and inclusion initiatives, including UBT?
4. Should a 'What Works' network be created for the equality, diversity and inclusion agenda, within which UBT will comprise one strand?

Organisations' ultimate purpose in implementing UBT interventions is to reduce or remove objective workplace inequalities. UBT is thought to be one method for achieving this aim. Gaps in the evidence prevent us from drawing robust conclusions about its impact beyond awareness raising and short-term implicit bias change. It is necessary to continue to accumulate and interrogate the evidence of UBT effectiveness as the adoption of UBT in UK businesses continues to increase.

Further research

The fundamental recommendation of this assessment is that more UK-based research and evaluations are required. The evidence directly comparing organisational approaches, design characteristics and outcomes for protected characteristics is very limited. Conclusions have been drawn based on single-approach interventions. The evidence would be immensely strengthened by systematic, controlled comparisons of the interventions considered in this assessment. This would allow for best-practice approaches to UBT to be identified and recommended.

Behavioural outcomes are currently limited to self-report assessments of behaviour or behavioural intentions. Research using valid measures of behaviour change is necessary to conclude the effectiveness of unconscious bias training for changing behaviour. Examples of valid measures include before-and-after measures of number of hires, involvement in diversity-relevant action or 360 degree feedback by the people who are subjected to bias.

To strengthen the evidence base regarding what works, further research should:

- Include systematic comparisons of approaches and design characteristics.
- Investigate whether UBT is equally effective for reducing bias against all protected groups.
- Ensure the valid measurement of outcomes of UBT, such as measures that assess actual behaviour change (as opposed to reported intentions to change).

- Assess the relative contribution of UBT and structural changes to achieving more equitable, diverse and inclusive organisational structures.
- Assess the relative contribution of other cognitive and social processes in maintaining inequity.
- Include real-life comparisons of strategies for managing and mitigating unconscious bias.
- Identify the boundary conditions that influence transfer learning from UBT into practice in the workplace.
- Examine the impact of mandatory versus voluntary attendance on UBT.
- Consider the circumstances under which UBT may have back-firing, or other counter-intuitive effects.

References

Atewologun, D., Cornish, T. and Briner, R. (2017), 'Responding to the reviews – an evidence-based approach to diversity & inclusion', *HR Magazine*, 26 June 2017. Available at: http://www.hrmagazine.co.uk/article-details/an-evidence-based-approach-to-diversity-and-inclusion [accessed: 28 February 2018].

Bezrukova, K., Spell, C.S., Perry, J.L. and Jehn, K. A. (2016), 'A meta-analytical integration of over 40 years of research on diversity training evaluation', *Psychological Bulletin*, vol. 142, no. 11, pp. 1227-74.

Business, Energy and Industrial Strategy (2017). 'Government response to Baroness McGregor Smith'. Avilable at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/594365/race-in-workplace-mcgregor-smith-review-response.pdf [accessed 8 March 2018].

Business in the Community (2012), *Opportunity Now: Excellence in practice awards: case studies 2012*. London: Business in the Community. Available at: http://www.globaldiversitypractice.co.uk/downloads-area/2012_Opps_Now_Case%20studies.pdf [accessed: 5 February 2018].

Business in the Community (2013), *Benchmarking trends analysis*. London: Business in the Community. Available at: https://race.bitc.org.uk/all-resources/research-articles/benchmarking-trends-analysis-2013 [accessed: 5 February 2018].

Burke, L.A. and Hutchins, H.M. (2007), 'Training transfer: an integrative literature review', *Human Resource Development Review*, vol. 6, no. 3, pp. 263-96.

*Capers IV, Q., Clinchot, D., McDougle, L. and Greenwald, A.G. (2017), 'Implicit racial bias in medical school admissions', *Academic Medicine*, vol. 92, no. 3, pp. 365-69.

Carnes, M., Devine, P.G., Isaac, C., Manwell, L.B., Ford, C.E., Byars-Winston, A. and Sheridan, J. (2012), 'Promoting institutional change through bias literacy', *Journal of Diversity in Higher Education*, vol. 5, no. 2, pp. 63-77.

*Carnes, M., Devine, P.G., Manwell, L.B., Byars-Winston, A., Fine, E., Ford, C.E., Forscher, P. Isaac, C. Kaatz, A., Maqua, W., Palta, M. and Sheridan, J. (2015), 'Effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial', *Academic Medicine*, vol. 90, no. 2, pp. 221-30.

Civil Service (2014), *How to do a REA*. London: Civil Service. Available at: http://webarchive.nationalarchives.gov.uk/20140402163101/http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/how-to-do-a-rea [accessed: 5 February 2018].

Correll, S.J. (2017), 'SWS 2016 Feminist Lecture: Reducing gender biases in modern workplaces: a small wins approach to organizational change', *Gender and Society,* vol. 31, no. 6, pp. 725-50.

Cornish, T. and Jones, P. (2013*), Unconscious bias in higher education: literature review*. London: Equality Challenge Unit. Available at: https://www.ecu.ac.uk/publications/unconscious-bias-in-higher-education/ [accessed: 5 February 2018].

*Devine, P.G., Forscher, P.S., Austin, A.J. and Cox, W.T.L. (2012), 'Long-term reduction in implicit race bias: a prejudice habit-breaking intervention', *Journal of Experimental Social Psychology*, vol. 48, no. 6, pp. 1267-78.

Dobbin, F., Schrage, D. and Kalev, A. (2015), 'Rage against the iron cage: the varied effects of bureaucratic personnel reforms on diversity', *American Sociological Review*, vol. 80, no. 5, pp. 1014-44.

Eagly, A H. and Karau, S.J. (2002), 'Role congruity theory of prejudice toward female leaders', *Psychological Review,* vol. 109, no. 3, pp. 573-98

*Forscher, P.S., Mitamura, C., Dix, E.L., Cox, W.T. and Devine, P.G. (2017), 'Breaking the prejudice habit: mechanisms, timecourse, and longevity', *Journal of Experimental Social Psychology,* vol. 72, pp. 133-46.

*Girod, S., Fassiotto, M., Grewal, D., Ku, M.C., Sriram, N., Nosek, B.A. and Valantine, H. (2016), 'Reducing implicit gender leadership bias in academic medicine with an educational intervention', *Academic Medicine*, vol. 91, no. 8, pp. 1143-50.

*Google. (2015), Guide: Raise Awareness about unconscious bias. Available at: https://rework.withgoogle.com/guides/unbiasing-raise-awareness/steps/watch-unconscious-bias-at-work/ [accessed: 5 March 2018].

*Hausmann, L.R., Long, J.A., Mor, M.K., Obrosky, D.S., Red, L.A., Rodriguez, C.J. and Saha, S. (2014), *Evaluation of a pilot program to improve patient health care experiences through PACT cultural competency training about unconscious bias*. Washington, DC: Department of Veterans Affairs, Veterans Health Administration.

Isaac, C., Lee, B. and Carnes, M. (2009), 'Interventions that affect gender bias in hiring: a systematic review', *Academic Medicine*, vol. 84, no. 10, pp. 1440-6.

*Jackson, S. M., Hillard, A.L. and Schneider, T.R. (2014), 'Using implicit bias training to improve attitudes toward women in STEM'. *Social Psychology of Education*, vol. 17, no. 3, pp. 419-38.

Kahneman, D., Lovallo, D. and Sibony, O. (2011), 'The big idea: before you make that big decision', *Harvard Business Review*, vol. 89, no. 6, pp. 50-60.

Kalev, A., Dobbin, F. and Kelly, E. (2006), 'Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies', *American Sociological Review*, vol. 71, no. 4, pp. 589-617.

Kawakami, K., Dovidio, J.F. and van Kamp, S. (2005) 'Kicking the habit: effects of nonstereotypic association training and correction processes on hiring decisions', *Journal of Experimental Social Psychology*, vol. 41, no. 1, pp. 68-75.

Kossek, E.E., Su, R. and Wu, L. (2017), '"Opting out" or "pushed out"? Integrating perspectives on women's career equality for gender inclusion and interventions', *Journal of Management*, vol. 43, no. 1, pp. 228-54.

Lai, C.K., Hoffman, K.M. and Nosek, B.A. (2013), 'Reducing implicit prejudice', *Social and Personality Psychology Compass*, vol. 7, no. 5, pp. 315-30.

*Lai, C.K., Marini, M., Lehr, S.A., Cerruti, C., Shin, J.E.L., Joy-Gaba, J.A., Ho, A.K., Teachman, B.A., Wojcik, S.P., Koleva, S.P., Frazier, R.S., Heiphetz, L., Chen, E.E.,

Turner, R.N., Haidt, J. Kesebir, S., Hawkins, C.B., Schaefer, H.S., Rubichi, S., Sartori, G., Dial, C.M., Sriram, N., Banaji, M.R. and Nosek, B.A. (2014), 'Reducing implicit racial preferences: I. A comparative investigation of 17 interventions', *Journal of Experimental Psychology: General*, vol. 143, no. 4, pp. 1765-85.

*Lai, C.K., Skinner, A.L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y.J., Pedram, C., Marshburn, C.K., Simon, S., Blanchar, J.C., Joy-Gaba, J.A., Conway, J., Redford, L., Klein, R.A., Roussos, G., Schellhaas, F.M.,Burns, M, Hu, X., McLean, M.C., Axt, J.R., Asgaris, S., Schmidt, K. Rubinstein, R., Marini, M., Rubichi, S., Shin, J.E. and Nosek, B.A. (2016), 'Reducing implicit racial preferences: II. Intervention effectiveness across time', *Journal of Experimental Psychology: General*, vol. 145, no. 8, pp. 1001-16.

Lee, C. (2017), *Awareness as a first step toward overcoming implicit bias*. GWU Law School Public Law Research Paper no. 2017-56. Available at: https://scholarship.law.gwu.edu/faculty_publications/1291/ [accessed: 5 March 2018].

*Leslie, K.F., Sawning, S., Shaw, M.A., Martin, L.J., Simpson, R.C., Stephens, J.E. and Jones, V.F. (2017), 'Changes in medical student implicit attitudes following a health equity curricular intervention', *Medical Teacher*. DOI: 10.1080/0142159X.2017.1403014 [online only].

*Lueke, A. and Gibson, B. (2015), 'Mindfulness meditation reduces implicit age and race bias: the role of reduced automaticity of responding', *Social Psychological and Personality Science*, vol. 6, no. 3, pp. 284-91.

*Lueke, A. and Gibson, B. (2016), 'Brief mindfulness meditation reduces discrimination', *Psychology of Consciousness: Theory, Research, and Practice*, vol. 3, no. 1, pp. 34-44.

McGregor-Smith, R. (2017), 'Race in the workplace': The McGregor-Smith review. London: Department for Business, Energy and Industrial Strategy. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/594336/race-in-workplace-mcgregor-smith-review.pdf [accessed: 5 February 2018].

Moss-Racusin, C.A., van der Toorn, J., Dovidio, J.F., Brescoll, V.L., Graham, M.J. and Handelsman, J. (2014), 'Scientific diversity interventions', *Science*, vol. 343, no. 6171, pp. 615-16.

*Moss-Racusin, C. A., van der Toorn, J., Dovidio, J. F., Brescoll, V. L., Graham, M. J. and Handelsman, J. (2016), 'A "scientific diversity" intervention to reduce gender bias in a sample of life scientists', *CBE-Life Sciences Education*, vol. 15, no. 3, article 29, pp. 1-11.

Nelson, O. (2017), 'Potential for progress: implicit bias training's journey to making change'. Undergraduate thesis, University of Pennsylvania.

Noon, M. (2018), 'Pointless diversity training: unconscious bias, new racism and agency', *Work, Employment and Society*, vol. 32, no. 1, pp. 198-209.

Oswald, F.L., Mitchell, G., Blanton, H., Jaccard, J. and Tetlock, P.E. (2013), 'Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies', *Journal of Personality and Social Psychology*, vol. 105, no. 2, pp. 171-92.

*PwC. (2016), *The PwC diversity journey*. Available at: https://www.pwc.com/gx/en/diversity-inclusion/best-practices/assets/the-pwc-diversity-journey.pdf [accessed: 5 February 2018].

*Repelaer van Driel, R. (2015), 'Women in science. The effect of training on gender bias reduction in academia.' Master's thesis, Institute of Psychology, Leiden University.

Rooth, D-O. (2010), 'Automatic associations and discrimination in hiring: real-world evidence', *Labour Economics,* vol. 17, no. 3: pp. 523-34.

Rudman, L.A. and Glick, P. (2001), 'Prescriptive gender stereotypes and backlash toward agentic women', *Journal of Social Issues*, vol. 57, no. 4, pp. 743-62.

Sherman, L.W., Strang, H. and Barnes, G.C. (1997), *The RISE Working Papers: the Canberra Reintegrative Shaming Experiments*. Canberra, Australia: Law Program, Research School for Social Sciences, Australian National University.

*Sweetman, J.P. (2017), *Evaluation of train the trainers unconscious bias training (Phase II)*. London: Equality Challenge Unit. Available at: https://ore.exeter.ac.uk/repository/handle/10871/30369 [accessed: 5 February 2018].

Wegener, D.T and Petty, R.E. (1997), 'The flexible correction model: the role of native theories of bias in bias correction'. In Zanna, M.P. (ed.) *Advances in experimental social psychology 29.* New York: Academic Press, pp. 141-208.

*Whatley, J.K. (2018), ‘Implicit bias as a contributing factor to disproportionality of African Americans in special education: the promise of a bias literacy intervention’. Ph.D. dissertation, Mercer University.

Wilson, T.D. and Brekke, N. (1994), ‘Mental contamination and mental correction: unwanted influences on judgments and evaluations’, *Psychological Bulletin*, vol. 116, no. 1, pp. 117-42.

References marked * were those deemed high quality evidence and were used to inform the conclusions drawn in this review.

Appendix 1|Methodology

A rapid evidence assessment methodology was used. This required a transparent and systematic approach to the search for evidence and the elimination of studies that did not meet pre-specified minimum quality standards.

It focused on one main assessment question:

- What is the evidence for the effectiveness of UBT?
    - Our focus was on 'effectiveness', how effectiveness is assessed (including indications of ineffectiveness), where it is effective, to what extent, and under what boundary conditions.

The authors agreed search terms, initial inclusion and exclusion criteria, databases to be searched and additional criteria for search results. A record was kept of search results, database sources and key decisions (for example the rationale for excluding/including certain literature). For primary research, we evaluated the robustness of the research methodology and conclusions.

To begin, we identified relevant and informative literature, which required two online searches. The first search aimed to collate academic literature, that is, peer reviewed journal articles. The second search aimed to collate non-academic literature, also known as 'grey' literature, that is, reports produced by private and public sector organisations.

We restricted search terms for both academic and non-academic literature to 'unconscious bias training' and 'implicit bias training' respectively. The searches for academic literature were conducted using Google Scholar. Academic literature was restricted to articles published since 2013. No geographical boundaries were placed on this search due to the limited amount of academic work published on unconscious bias training in the UK. The searches for the grey literature were conducted using the Google search engine. Grey literature was restricted to org.uk, gov.uk and ac.uk domains and only PDF file formats to reduce irrelevant hits.

Criteria for inclusion

The initial online searches returned 2,701 sources. Each source was reviewed to assess its relevance. Academic sources were included if they provided either a) an outcome or evaluation of one or more unconscious bias training interventions or programmes or b) a theoretical argument, debate or review relating to unconscious bias training that provided information for or against why, and under what circumstances, unconscious bias training may be effective. Grey literature sources were included if they evaluated or provided outcomes for unconscious bias training.

The final sample of sources deemed relevant for the assessment was 88 (comprising 57 academic sources and 31 grey literature sources). These 88 sources were then evaluated.

Evaluation of evidence

To assess the methodological quality of the evidence, we used the Maryland Scale of Scientific Methods (MSSM) as recommended by the Civil Service in the Rapid Evidence Assessment Toolkit (Civil Service, 2014).

The MSSM was developed by Sherman *et al*. (1997). The scale of 1 to 5 (based on the robustness of the method used), gives us an established framework to assess quality of the available evidence for the effectiveness of unconscious bias training. The scale is outlined in Table A1 (below).

Table A1.1 Description of levels in the Maryland Scale of Scientific Methods

Level 1	Either: a) A correlation between a training intervention and outcomes at a single point in time. For example: A case study that reported an increase in women's progression after the implementation of UBT. b) End-of-intervention evaluation. For example: A study that measured the impact of the UBT workshop on participants using a questionnaire at the end of the event.
Level 2	Either: a) Temporal sequence between the training intervention and the outcome clearly observed.

	For example: A study that measured training participants' UB test scores before and after a training intervention. b) The use of a comparison group that has not been 'matched' or selected based on its similarity to the test group. For example: A study that compared UB test scores after a training intervention between training participants and everybody else in the same organisation.
Level 3	A comparison between two or more comparable units of analysis, one with and one without the intervention. For example: A study that compared UB test scores before and after a training intervention between training participants and a matched group from the same organisation, where the matched group was similar to the training group in factors such as gender, ethnicity, tenure and functional mix.
Level 4	A comparison between multiple units with and without the training intervention, controlling for other factors or using comparison units that evidence only minor differences. For example: A study that compared UB test scores before and after a training intervention between training participants and a matched group from the same organisation, where the matched group is similar to the training group in gender, ethnicity, tenure and functional mix. In addition, when analysing scores, statistical techniques were used to ensure that the programme and comparison groups were as similar as possible, by controlling for any effects of spurious factors. This statistical control increases confidence that any observed differences in scores are due to the different conditions, rather than other factors.
Level 5	Random assignment and analysis of comparable units to intervention and control groups. For example: A study that started with randomly assigning people into training and matched groups, then compared both groups on UB test scores before and after the training intervention. It used

	statistical techniques to check that the programme and comparison groups were similar, by controlling for the effects of spurious factors during analysis. This statistical control increases confidence that any observed differences in scores is due to the different conditions, rather than other factors.

For inclusion in this assessment, we selected studies that reached a minimum level of 2 according to the MSSM. The final list of studies used for this assessment, and the evidence against which our recommendations are based, is listed in Table A1.2 (below).

Table A1.2 Sources of evidence used in this assessment

Source	MSSM Rating	Aims of UBT			
		Awareness Raising	Implicit Bias Change	Explicit Bias Change	Behaviour Change
Capers *et al*. (2017)	2	✓			✓
Moss-Racusin *et al*. (2016)	2	✓		✓	✓
Girod *et al*. (2016)	2	✓	✓	✓	

PwC (2016)	2	✓			✓
Whatley (2018)	2	✓			✓
Leslie *et al*. (2017)	2		✓		
Sweetman (2017)	2	✓	✓	✓	✓
Google (2013)	3	✓			✓
Repelaer van Driel (2015)	4				✓
Lai *et al*. (2014)	4		✓	✓	
Lai *et al*. (2016)	4		✓	✓	

Hausmann *et al*. (2014)	4	✓		✓	✓
Devine *et al*. (2012)	4	✓	✓	✓	
Forscher *et al*. (2017)	4	✓	✓	✓	✓
Lueke and Gibson (2015)	5		✓		
Lueke and Gibson (2016)	5		✓		
Jackson *et al*. (2014)	5		✓	✓	
Carnes *et al*. (2015)	5	✓	✓		✓

Definition of sources

Below are descriptions of the categories of sources used in this assessment.

Academic interventions:

- Design and implement an intervention based on theoretical arguments or previous research
- Implement and evaluate the intervention using scientific methods
- Report evidence-based outcomes of the intervention
- Are published in a peer-reviewed journal

Academic reviews:

- Provide a (or multiple) theoretical framework(s)
- Often report on several (up to hundreds of) studies
- Bring a new or insightful argument to the discussion/ debate
- Are published in a peer-reviewed journal

Academic research studies:

- Provide a research question or topic of investigation
- Adopt scientific research methods to test hypotheses
- Report evidence-based outcomes of the research
- Are published in a peer-reviewed journal

Reports and non-academic research:

- Have a clearly defined topic
- Provide outcomes, conclusions or recommendations that are evidence-based
- Are published by a professional body (such as the UK Government, a charity or private organisation)

Case studies:

- Have a clearly defined aim, topic or intervention
- Are about a specific individual, group or organisation

Appendix 2|Additional Sources

The following additional sources were deemed relevant but were not used as the evidence base for the assessment for one (or both) of the following reasons:

- The source provides a theoretical argument but does not provide 'evidence' from research.
- The quality of the evidence provided in the source was not deemed sufficient to be included.

The sources have been provided to give an insight into current debates, thinking and/or practices for designing, implementing or evaluating unconscious bias training.

Akram, S. (2017), 'Representative bureaucracy and unconscious bias: Exploring the unconscious dimension of active representation', *Public Administration*. DOI: 10.1111/padm.12376 [online only]

Allen, B.J. and Garg, K. (2016), 'Diversity matters in academic radiology: acknowledging and addressing unconscious bias', *Journal of the American College of Radiology*, vol. 13, no. 12, pp. 1426-32.

Ashworth, A. (2014), *Educating staff about lesbian, gay and bisexual equality*. London; Stonewall. Available at: https://www.stonewall.org.uk/sites/default/files/training_guide.pdf [accessed: 5 February 2018].

Beech, N., Cornelius, N., Gordon, L., Healy, G, Ogbonna, E., Sanghera, G., Umeh, C., Wallace, J. and Woodman, P. (2017), *Delivering diversity. Race and ethnicity in the management pipeline*. London: Chartered Management Institute and British Academy of Management. Available at: http://www.managers.org.uk/~/media/Files/PDF/Insights/CMI_BAM_Delivering_Diversity_2017_Full_Report_Website_Copy.pdf [accessed: 5 February 2018].

Bennett, M. and Plaut, V. (2018), 'Looking criminal and the presumption of dangerousness: Afrocentric facial features, skin tone, and criminal justice', *UC Davis Law Review*, vol. 51, no. 3, pp. 745-803.

Berk, R.A. (2017), 'Microaggressions trilogy: part 2. Microaggressions in the academic workplace', *The Journal of Faculty Development*, vol. 31, no. 2, pp. 69-83.

Birckhead, T.R. (2017), 'The racialization of juvenile justice and the role of the defense attorney', *Boston College Law Review*, vol. 58, no. 2, pp. 379-461.

Birdi, K. (2010). *The taxonomy of training and development outcomes (TOTADO) evaluation framework*. The Management School, University of Sheffield. Available at: https://www.researchgate.net/profile/Kamal_Birdi/publication/273699650_The_Taxonomy_of_Training_and_Development_Outcomes_TOTADO_Evaluation_Framework/links/55097c850cf2d7a2812cd10a/The-Taxonomy-of-Training-and-Development-Outcomes-TOTADO-Evaluation-Framework [accessed: 5 February 2018].

Business in the Community, Opportunity Now (2014), *Excellence in practice awards: case studies 2014*. London: Business in the Community. Available at: https://gender.bitc.org.uk/awards-benchmarking/excellence-practice-awards-showcase-2014 [accessed: 5 February 2018].

Business in the Community (2015), *Race at work report*. London: Business in the Community. Available at: https://race.bitc.org.uk/all-resources/research-articles/race-work-report [accessed: 5 February 2018].

Business in the Community (2015), *Sector factsheet: ethnic minorities in science, technology, engineering and mathematics (STEM).* London: Business in the Community. Available at: https://race.bitc.org.uk/sites/default/files/rfo_sector_factsheet_set_vfinal_new_2.pdf [accessed: 5 February 2018].

Business in the Community (2016), *Business in the Community workplace gender equality awards 2016*. London: Business in the Community. Available at: https://gender.bitc.org.uk/sites/default/files/gender_awards_2016_case_studies.pdf [accessed: 5 February 2018].

Bond, S. and Wollaston, H. (2015), *Routes to diversity and inclusion toolkit*. London: Chartered Institution of Highways & Transportation. Available at: http://www.ciht.org.uk/en/document-summary/index.cfm/docid/C83B8131-5FB7-49CE-BE0BBDAEA8889550 [accessed: 5 April 2018].

Boscardin, C.K. (2015), 'Reducing implicit bias through curricular interventions', *Journal of General Internal Medicine*, vol. 30, no. 12, pp. 1726-28.

Brown, D.A. (2014), 'Diversity and the high-tech industry'. *Alabama Civil Rights and Civil Liberties Law Review*, vol. 6, pp. 95-123.

Care Quality Commission (2016), *The Christie NHS Foundation Trust: quality report*. London: Care Quality Commission. Available at: http://www.cqc.org.uk/sites/default/files/new_reports/AAAF8405.pdf [accessed: 5 February 2018].

Civil Service (2015), *Identifying and removing barriers to talented BAME staff progression in the Civil Service*. London: Civil Service. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/417250/Ethnic_Dimension_Blockages_to_Talented_BAME_staff_Progression_in_the_Civil_Service_Final_16.12.14__1_.pdf [accessed: 5 February 2018]

Civil Service (2017), *A brilliant Civil Service: becoming the UK's most inclusive employer*. London: Civil Service. Available at: https://quarterly.blog.gov.uk/wp-content/uploads/sites/5/2017/11/a_brilliant_civil_service_becoming_the_uks_most_inclusive_employer.pdf [accessed: 5 February 2018]

Day, M. (2017), 'Facilitating a positive culture for the disclosure of workers disabilities: experiences of developing a pilot volunteer intervention programme within a higher education institution', *Journal of Healthcare Communications*, vol. 2, no. 1.

Digital Scotland and Equate Scotland. (2017), *Tackling the technology gender gap together*. Edinburgh; Equate Scotland. Available at: http://www.equatescotland.org.uk/wp-content/uploads/2017/04/bestpracticeguide.pdf [accessed: 5 February 2018].

Dorio, J.K. (2015), 'Cognitive bias and cultural responsiveness: a training course for the judiciary'. Ph.D. dissertation, Azusa Pacific University.

Equate Scotland. (2017), *Women in engineering: employer's best practice guide*. Edinburgh: Equate Scotland. Available at: http://www.equatescotland.org.uk/wp-content/uploads/2016/06/Women-in-Engineering-Employers-Best-Practice-Guide.pdf [accessed: 5 February 2018].

Febbraro, A.R. and Pickering, D.I. (2015), *Women in Science, Technology, Engineering, Mathematics, and Management*. Scientific Report DRDC-RDDC-2015-

R187, Toronto: Defence Research and Development Canada. Available at: http://cradpdf.drdc-rddc.gc.ca/PDFS/unc258/p802775_A1b.pdf [accessed: 5 March 2018].

Gale, A. (2017), 'Examining Black adolescents' perceptions of in-school racial discrimination: the role of gender, socioeconomic status and perceptions of teacher support on academic outcomes.' Ph.D. dissertation: University of Michigan. Available at: https://deepblue.lib.umich.edu/handle/2027.42/138696 [accessed: 5 March 2018].

Giacomantonio, C., Jonathan-Zamir, T., Limanovitz, Y., Bradford, B., Davies, M., Strang, L. and Sutherland, A. (2016), *College of Policing Stop and Search Training Experiment: Process Evaluation*. RAND Corporation: College of Policing Limited. Available at: https://www.rand.org/pubs/research_reports/RR1442.html accessed: 5 March 2018].

Gullo, G L. (2017), 'Implicit bias in school disciplinary decisions'. Doctoral dissertation, Lehigh University.

Handelsman, J. and Sakraney, N. (2015), *Implicit bias*. White House Office of Science and Technology Policy. Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/bias_9-14-15_final.pdf [accessed: 5 March 2018].

Harris, C. (2017), 'Dark innocence: retraining police with mindfulness practices to aid in squelching implicit bias', *University of San Francisco Law Review*, vol. 51, no. 1, pp. 103-25.

Her Majesty's Inspectorate of Constabulary and Fire & Rescue Services. (2017), *PEEL: police legitimacy (including leadership) 2017: An inspection of City of London Police*. Available at: https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/peel-police-legitimacy-2017-city-of-london.pdf [accessed: 5 February 2018].

Her Majesty's Inspectorate of Constabulary and Fire & Rescue Services (2017), *PEEL: police legitimacy (including leadership) 2017: an inspection of Northamptonshire Police*. London: HMICFRS. Available at: https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/peel-police-legitimacy-2017-northamptonshire.pdf [accessed: 5 February 2018].

HM Revenue & Customs (2015), *Equality Act 2010 (Specific Duty) regulations 2011, section 149: How HMRC complies with its public sector equality duties*. London: HMRC. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/400088/Equality_Act_2010__Specific_Duty__Regulations_2011_section_149_-_How_HMRC_complies_with_its_public_sector_equality_duties..pdf [accessed: 5 February 2018].

Jackson, J.L. (2017), 'Uniform feelings: emotion management and police work '. Ph.D. dissertation, State University of New York at Buffalo.

James, L. (2018), 'The stability of implicit racial bias in police officers', *Police Quarterly*, vol. 21, no. 1, pp. 30-52.

Jones, T. and Norwood, K.J. (2016), 'Aggressive encounters and White fragility: deconstructing the trope of the angry Black woman', *Iowa Law Review*, vol. 102, pp. 2017-69.

Kamerlin, S.C.L. (2016), 'Where are the female science professors? A personal perspective', *F1000Research*, no: 5: 1224, pp. 1-17.

Kofman, L. and Mather, L. (2017), 'The bias of crowds model: promise and potential challenges for practitioners', *Psychological Inquiry*, vol. 28, no. 4, pp. 278-80.

Kulik, C.T. and Roberson, L. (2008), 'Common goals and golden opportunities: Evaluations of diversity education in academic and organizational settings', *Academy of Management Learning and Education*, vol. 7, no. 3, pp. 309-31.

Law Society (2013), *Diversity and inclusion charter annual review 2013*. London: The Law Society. Available at: https://www.google.co.uk/url?sa=tandrct=jandq=andesrc=sandsource=webandcd=2andved=0ahUKEwihs5qn8Y7ZAhUEI1AKHY5nDdUQFggxMAEandurl=http%3A%2F%2Fwww.lawsociety.org.uk%2Fsupport-services%2Fpractice-management%2FDiversity-inclusion%2Fdocuments%2FDiversity-and-inclusion-annual-report--December-2013andusg=AOvVaw2pK8r7OiB5mormGR9h7D8D [accessed: 5 February 2018]

Levy, K. and Barocas, S. (2018), 'Designing against discrimination in online markets', *Berkeley Technology Law Journal*, vol. 32 (forthcoming). Available at:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3084502 [accessed: 5 March 2018].

Miller, D.A. (2017), *Physicians hold implicit* biases *about women with cervical cancer*. University of Arizona. Available at: http://arizona.openrepository.com/arizona/handle/10150/625103 [accessed: 5 March 2018].

McCormick, H. (2015), *The real effects of unconscious bias in the workplace.* UNC Executive Development, Kenan-Flagler Business School, University of North Carolina. Available at: https://coup-usa.com/FILES/THE%20EVIDENCE/unc-white-paper-the-real-effects-of-unconscious-bias-in-the-workplace-Final.pdf [accessed: 5 February 2018].

Michael, L. (2016), *Race equality works for Northern Ireland*. Belfast: Northern Ireland Relations Council. Available at: http://www.bitcni.org.uk/wp-content/uploads/2016/10/Race-Equality-Works-for-Northern-Ireland-2016.pdf [accessed: 5 February 2018].

Mitchell, G. (2018), 'Jumping to conclusions: advocacy and application of psychological research'. In Crawford, J.T. and Jussim, L. (eds.) *The politics of social psychology*. New York: Routledge (forthcoming).

Odeh, A., Bruce, T.J., Krenn, D.R. and Ran, S. (2017), 'A broader perspective for subtle discrimination interventions', *Industrial and Organizational Psychology*, vol. 10, no. 1, pp. 118-23.

Paluck, E.L. and Green, D.P. (2009), 'Prejudice reduction: What works? A review and assessment of research and practice', *Annual Review of Psychology*, vol. 60, pp. 339-67.

Payne, B.K. and Vuletich, H.A. (2018), 'Policy insights from advances in implicit bias research'. *Policy Insights from the Behavioral and Brain Sciences*, vol. 5, no. 1, pp. 49-56.

Payne, B.K., Vuletich, H.A. and Lundberg, K. B. (2017), 'The bias of crowds: How implicit bias bridges personal and systemic prejudice', *Psychological Inquiry*, vol. 28, no. 4, pp. 233-48.

Perez, N.M., Sim, S., Turner, A. and Argueta, S. (2013), 'Launching a race equity project in the City of Angels', *Clearinghouse Review: Journal of Poverty Law and Policy*, vol. 47, no. 5-6, pp. 250-55.

Randall, V., and Randall, T. (2016), 'Cutting across the bias: teaching implicit bias in a healthcare law course', *Saint Louis University Law Journal*, vol. 61, pp. 511-28.
Royal Academy of Engineers (2015)*, Increasing diversity and inclusion in engineering- a case study toolkit*. Available at: https://www.raeng.org.uk/policy/diversity-in-engineering/diversity-and-inclusion-toolkit/documents/increasing-diversity-and-inclusion-in-engineering [accessed: 5 February 2018].

Shabbar, A. (2017), 'Promoting inclusion through organizational culture change'. Master of Design dissertation, Toronto. Available at: http://openresearch.ocadu.ca/id/eprint/1833/ [accessed: 5 March 2018].

Smith, R.J. (2015), 'Reducing racially disparate policing outcomes: is implicit bias training the answer', *University of Hawaii Law Review*, vol. *37,* pp. 295-312.

Spicer, A. and Culpin, V. (2017), 'Unleashing the power of diversity: measurement, motivation and management.' Paper for Power and Diversity Forum, May. Available at: https://www.cityoflondon.gov.uk/business/economic-research-and-information/research-publications/Documents/research-2017/Unleashing-the-Power-of-Diversity-v2.pdf [accessed: 5 February 2018].

Taddia, M. (2013), *A diverse workforce that understands its diverse clients*. London: International Property Regulation Board Available at: https://ipreg.org.uk/wp-content/files/2013/04/Article1.pdf [accessed: 5 February 2018].

Tsai, J.W. and Michelson, C.D. (2017), 'Implicit bias training in pediatric residency: attitudes amongst program directors and lessons learned from implementation (Research abstract). *Academic Pediatrics*, vol.17, no. 5, pp. e53-e54.

Vinnicombe, S., Doldor, E. and Turner, C. (2014), *The female FTSE board report 2014: crossing the finish line.* Cranfield: Cranfield School of Management. Available at: https://dspace.lib.cranfield.ac.uk/bitstream/handle/1826/8365/Female_FTSE_Report_2014.pdf;jsessionid=53F4837166619D7DDBBFB65E15CBF997?sequence=1 [accessed: 5 February 2018].

Warmington, J. and Kline, R. (2015), *Unconscious bias: silver bullet or just a useful tool?* Available at: https://www.linkedin.com/pulse/unconscious-bias-silver-bullet-just-useful-tool-joy-warmington/ accessed: 5 February 2018].

West, S. (2016), *Global allies: engage, educate, empower*. London: Stonewall. Available at: https://www.stonewall.org.uk/sites/default/files/global_allies_-_engage_educate_empower.pdf [accessed: 5 February 2018].

Wistrich, A.J. and Rachlinski, J.J. (2017), *Implicit bias in judicial decision making how it affects judgment and what judges can do about it*. In American Bar Association, *Enhancing justice*. Cornell Legal Studies Research Paper no. 17-16. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2934295 [accessed: 5 March 2018].

Wright, A. (2016), *Police interactions with individuals with developmental disabilities: use of force, training, and implicit bias*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903331 [accessed: 5 March 2018].

Yost, B.S. (2017), 'What's wrong with differential punishment?' *Utilitas*, vol. 29, no. 3, pp. 257-85.

Contacts

This publication and related equality and human rights resources are available from the Commission's website: www.equalityhumanrights.com.

Questions and comments regarding this publication may be addressed to: correspondence@equalityhumanrights.com. The Commission welcomes your feedback.

Alternative formats

This publication is also available as a Microsoft Word file from www.equalityhumanrights.com. For information on accessing a Commission publication in an alternative format, please contact: correspondence@equalityhumanrights.com.

EASS

For advice, information or guidance on equality, discrimination or human rights issues, please contact the Equality Advisory and Support Service, a free and independent service.

Website	www.equalityadvisoryservice.com
Telephone	0808 800 0082
Textphone	0808 800 0084
Hours	09:00 to 19:00 (Monday to Friday) 10:00 to 14:00 (Saturday)
Post	FREEPOST EASS HELPLINE FPN6521

Published March 2018

You can download this publication from

www.equalityhumanrights.com

Published: March 2018

Equality and
Human Rights
Commission