LUNDS UNIVERSITET

Department of Psychology

ACT to Facilitate Change – Can Conversation about Injustice in a Theatre Setting Reduce Inflexibility, Prejudice and Negative Reactions Towards Discussions of Discrimination?

Elise Brinck

Master's Thesis (30hp)
Spring 2024

Supervisors: Sima Nurali Wolgast & Eva Hoff

Abstract

This study aimed to explore the effect of group conversation about injustice using methods from Acceptance and Commitment Training (ACTr) in a theatre setting. The objective was to reduce prejudice, psychological inflexibility with stigma and white fragility as well as increase compassion and behavioural intentions towards equality, among a sample of Swedish adults. An interactive conversational theatre was used as the intervention condition and was compared to a control group. One session was held, where participants either took part in the conversational theatre or a focus discussion group. A total of 85 participants took part in the study, which was carried out in Malmö, Sweden. The findings indicated no effects on compassion or behavioural intentions towards equality in either group. Prejudice was observed to increase in both groups, while white fragility increased in the intervention but not the control group. Psychological inflexibility was observed to decrease in both groups, but the intervention was significantly more effective. The results contradict some previous findings and future research is needed to further elucidate these effects. Outcomes and implications are discussed in the context of prejudice reduction interventions, entertainment narrative approaches, and Acceptance and Commitment Theory.

Keywords: prejudice reduction, compassion, psychological inflexibility, acceptance and commitment training, narrative persuasion, theatre

Acknowledgements

I want to extend my deepest thanks to everyone who has contributed to this project and helped it come to fruition. First and foremost, thank you to everyone who participated in the study, for giving me your valuable time and your eager engagement. Without your participation, this thesis could not exist. I also want to thank my supervisors Sima Nurali Wolgast and Eva Hoff, for all your help, dedication, and expert advice. A big thank you to Jörgen Dahlqvist, for your instrumental contribution to the creation and execution of this project. Finally, a warm thank you to my friends and family for your continuous support and encouragement.

Introduction

Discrimination of ethnic minorities is a significant social issue in Sweden (Diskrimineringsombudsmannen [DO], 2022). In recent years, there has been an increase in reports of discrimination based on race and/or ethnicity (DO, 2022). Moreover, a recently published report established that there are significant racial inequalities in the job market in Sweden, and that many non-white Swedes experience racism and discrimination in the workplace (Wolgast & Wolgast, 2021). For example, non-white Swedes are overall less satisfied with their jobs and are significantly more likely to express intentions to quit than those who are white (Wolgast & Wolgast, 2021). Taken together, it is evident that issues of discrimination and racism faced by non-white people both on the job market as well as in other contexts in Sweden are significant.

Beyond the Swedish job market, research has extensively documented the prevalence and adverse consequences of racism and discrimination in various contexts (e.g., Dover et al., 2020; Dovidio et al., 2017; Pager & Shepherd, 2008). For example, discrimination has been found to have substantial harmful consequences for the health and well-being of those who are subjected to it (Dover et al., 2020). This is proposed to be due to (a) direct effects of unequal resource distribution and quality of healthcare available, (b) stress-related physiological effects of encountering discrimination, and (c) behavioural responses and coping strategies in response to facing discrimination that undermine health (Dover et al., 2020). Thus, it can be regarded as a public health issue to make efforts to reduce discrimination.

Negative stereotypes contribute to the perpetuation of discrimination, shaping attitudes and behaviours toward marginalized groups (Devine et al., 2012). We also often see resistance to discussions of injustice, which can stem from individuals' discomfort with confronting systemic inequalities or acknowledging their complicity (DiAngelo, 2011; Hill et al, 2021). In the past, interventions targeting negative stereotypes and resistance to equality have predominantly focused on educational approaches, such as diversity training and various prejudice reduction programs (Paluck & Green, 2009; Paluck et al., 2021). While these interventions have shown some effectiveness, their impact often remains limited and short-lived due to factors such as reactance effects (i.e., negative backlash to the intervention) and/or lack of active engagement (Paluck et al., 2021). The persistence of discrimination highlights a critical societal issue. Proactive intervention and mitigation strategies are therefore essential, and ongoing work is necessary to develop truly effective solutions with lasting effects. The present project endeavours to investigate the effect of a group-based

conversation about injustice intervention aimed at addressing racial prejudice and resistance towards discussions of racism and injustice. By employing an interactive group format, based on methods from third-wave behavioural therapies[1], and testing this format in a novel setting inspired by conversational theatre performance, this study aims to explore an active method for reducing prejudice. Specifically, the intervention is geared towards increasing compassion, confronting prejudice, reducing inflexibility with stigma and combating resistance to discussing discrimination.

Discrimination, prejudice and 'modern racism'

Within the scope of this project, several terms such as 'prejudice', 'stereotypes', 'discrimination' and 'racism' are used. These terms can have different descriptions, and thus definitions are initially discussed. 'Prejudice' is defined as "a preconceived negative attitude towards certain individuals or groups, which have sometimes formed without any prior interaction with said individual or group" (APA Dictionary of Psychology, 2024a). Prejudice may result in emotional responses such as anxiety, anger, disdain, or even hatred, alongside cognitive responses such as presumptions and group-based beliefs, commonly known as 'stereotypes' (APA Dictionary of Psychology, 2024a). 'Discrimination' and discriminatory behaviour is defined as the outward expression of prejudice (APA Dictionary of Psychology, 2024b). Finally, the definition of 'racism' has varied over time, but a common denominator among definitions is that it involves the devaluation or denial of a person's value and abilities based on their ethnicity, skin colour and/or culture. In the context of this thesis, racism is used to refer to any form of discrimination based on race, ethnicity and/or skin colour. Discrimination is often considered to be one of the worst consequences of racism due to the many adverse effects that it has both on an individual and a systemic level for those who are subjected to it (Dover et al., 2020).

Within modern Western society, there exists a general understanding of discrimination and racism as negative and socially unacceptable. However, expressions of racism have, according to some scholars, not necessarily been reduced but have rather changed their form (Akrami et al., 2000; Dovidio et al., 2008; McConahay, 1986; Pettigrew & Meertens, 1995**).** Throughout history, ideas behind 'classic racism' such as racial biology (also referred to as

[1] Forms of Cognitive and Behavioural Therapies (CBTs) emphasising mindfulness, emotions, acceptance, values, goals, and meta-cognition (Hayes & Hoffman, 2017)

'scientific racism')[2] have been argued to justify overtly unequal treatment of and discrimination against people of ethnicities that are perceived as 'inferior' (Golec de Zavala & Cichocka, 2012). Nowadays, overt discrimination and expressions of racism are generally not socially acceptable. However, racism arguably still exists, albeit in a new form. The literature describes several types of 'new racism' such as *modern racism, subtle prejudice,* and *aversive racism,* all of which are assumed to be disguised and more covert forms of classical overt racism (Akrami et al., 2000). 'Modern racism' as coined by McConahay (1986) posits that white people still harbour negative feelings towards people of other ethnicities, but express them in different ways than outright aggression. The theory of 'subtle prejudice' suggests that racism is a manifestation of three components: (1) the defence of traditional values such as the importance of a high work ethic, (2) an exaggeration of cultural differences, which is often expressed through stereotyping, and (3) the denial of positive feelings about the out-group as a way of compensating for the fact that it is no longer acceptable to show overt negative feelings (Pettigrew & Meertens, 1995). 'Aversive racism' is a theory that postulates that white people can genuinely believe that equality is important, and that prejudice and discrimination are reprehensible, while still harbouring unconscious negative feelings towards non-whites that are based on fear (Gaertner & Dovidio, 2005). These conflicting feelings create anxiety, which in turn leads to avoidance of contact with out-groups (Dovidio et al., 2008; Gaertner & Dovidio, 2005). All above mentioned theories taken together paint a picture of 'modern racism' where it is clear that non-white people still face discrimination in many different ways, such as denial of their experiences, stereotyping and exclusion.

White fragility and counter-reactions to discussions of Racism

One expression of the aforementioned 'modern racism' is *white fragility*. This term refers to a concept which was first defined by Robin DiAngelo (2011), to describe the reactions and emotional discomfort exhibited by some white individuals when they are confronted with discussions regarding race and racism. These reactions often stem from uneasiness with acknowledging the existence of systemic racism, the existence of white privilege, as well as the ways in which white privilege operates in society (DiAngelo, 2011). White privilege, in turn, refers to the privileges that white people have in our current society due to being white, such as not having to face or worry about discrimination (DiAngelo, 2011). When confronted with issues of inequality or racism, research has found that

[2] An ideology insisting that people can be divided into races based on skin colour and ethnicity, and that whites as a race are superior to others (Harvard Library, 2024)

individuals displaying white fragility may become defensive, hostile, or dismissive (DiAngelo, 2011; Hill et al, 2021). Furthermore, they may feel personally attacked or offended by discussions that challenge their understanding of race, or suggestions that they have benefited from certain racial privileges (DiAngelo, 2011; Hill et al, 2021). This defensiveness often arises from a fear of being seen as racist, or from a discomfort with acknowledging one's own racial biases and prejudices (DiAngelo, 2011). These kinds of reactions are also often found within research on interventions to reduce racism, and counter-reactions to such efforts are not uncommon (Hill et al., 2021; Paluck, 2006; Plaut et al., 2011). However, white fragility does not only manifest as hostile or defensive counter-reactions to discussions of racism. It can also manifest as (a) a tendency to minimize or deny the experiences of people of colour, (b) deflect blame onto individual actions rather than systemic issues, and (c) as avoidance towards engaging in conversations about race altogether (DiAngelo, 2011; Hill et al, 2021). This also relates to the theory of aversive racism (Dovidio et al., 2008; Gaertner & Dovidio, 2005). In essence, white fragility reflects a reluctance to confront uncomfortable truths about racism and its impact on society (DiAngelo, 2011). Research also suggests that the behavioural manifestations of white fragility may be more common in specific contexts, such as diversity training seminars (Hill et al., 2021).

Displays of white fragility are further exemplified in the previously mentioned report regarding racism and discrimination in the job market in Sweden (Wolgast & Wolgast, 2021). This report illustrates that it is common for non-white people who raise issues or point out injustices related to ethnicity and/or skin colour in their workplace to be met with negative counter-reactions from their white colleagues (Wolgast & Wolgast, 2021). Moreover, a similar picture emerged in the responses from managers and supervisors. A large proportion of people stated that there existed resistance within their organization to working with equality issues, because it may cause "white Swedes" to feel accused, blamed and unfairly treated (Wolgast & Wolgast, 2021). Taking this into account, it is evident that expressions of white fragility are part of the racial discrimination that non-white people in Sweden face in the job market. Moreover, this also illustrates that recognizing white fragility, and the ways it manifests, is crucial for understanding how to build interventions aimed at reducing prejudice and/or discrimination. Finding ways to reduce white fragility, and combat people’s avoidance of discussing the topic of inequality, are key things to focus on in the context of interventions aimed at reducing discrimination and racism.

Methods for reducing discrimination – intergroup contact, perspective-taking and compassion

Actions against racism and discrimination often focus on reducing prejudice (Plauck & Green, 2009; Paluck et al., 2021). As previously mentioned, discrimination often serves as an outward expression of prejudice or negative attitudes toward certain individuals or groups. However, reducing prejudice is not an easy task. Attempts to reduce prejudice in natural settings have often been less successful than in laboratory studies (Paluck et al., 2021). The most empirically researched approach for reducing prejudice and discrimination is Allport's (1954) *contact hypothesis*. This approach is based on the idea that prejudice can be reduced by allowing people from different groups to interact with each other under favourable conditions, such as having a positive conversation or engaging in a fun activity. Recent studies also support the effectiveness of the contact hypothesis in modern times (Lowe, 2020). However, contact between members of different groups works best if its facilitating conditions are met (Allport, 1954). These conditions are: equal group status within the contact situation, common goals, intergroup cooperation, and support from authority (Allport, 1954; Wright et al., 1997). However, direct contact is often not possible or feasible in real life, due to issues of segregation and deep-rooted inequality. Thus, indirect contact, which does not involve face-to-face interaction, is often used in contemporary research (White et al., 2021). Direct and indirect contact have both been found to reduce prejudice (Pettigrew et al., 2007). Moreover, indirect contact could be seen as a first step towards direct contact, since it has been found to reduce intergroup anxiety (i.e., anxiety regarding interacting with members of an outgroup; Wölfer et al., 2019).

Beyond the contact hypothesis, other approaches and hypotheses for how prejudice and discrimination can be reduced have also been proposed. One such approach is ‘consciousness-raising’ methods, which aim to reduce prejudice by questioning the accuracy of stereotypes and prejudices, as well as emphasizing the negative consequences of prejudice (Hsieh et al., 2022). An example is the ‘self-regulation of prejudice model’, which usually involves making people aware of discrepancies between their prejudices and core values (Monteith et al., 2016). The idea is that the desire to act in line with one's core values can motivate a reduction in prejudice if discrepancies are pointed out (Monteith et al., 2016). Furthermore, there exists some initial support for the effectiveness of prejudice reduction interventions grounded in mindfulness, which have proved to increase people’s awareness of negative stereotypes as well as their openness and acceptance of differences among people (Fuochi et al., 2023).

Other important aspects of prejudice reduction are perspective-taking and compassion. Perspective-taking can have a positive effect on outgroup evaluations (Shih et al., 2009), and lead to a reduction in stereotyping and prejudice (Vescio et al., 2003). Increased perspective-taking has also been found to reduce automatic expressions of bias (Todd et al., 2011). In addition, feelings of compassion[3] and empathy[4] have been found to be associated with reduced expressions of prejudice (Berger et al., 2018). Thus, there appears to be a noteworthy connection between perspective-taking, compassion, and prejudice. These relationships warrant further investigation. Namely, investigating how perspective-taking and compassion could be activated or evoked in different ways is an interesting avenue for research on prejudice reduction.

Narrative persuasion – using entertainment, media and narratives to reduce prejudice and stigma

As previously established, the positive effects of intergroup contact can occur not only through direct in-person contact but also through indirect contact (Pettigrew et al., 2007; White et al., 2021; Wölfer et al., 2019). One such form of indirect contact is entertainment media, and in particular media which portrays personal narratives (Park, 2012). 'Mediated intergroup contact', defined as the connection between an ingroup observer and an outgroup narrative protagonist, has been shown to effectively reduce prejudice and improve attitudes toward marginalized groups (Igartua & Frutos, 2017; Igartua et al., 2019; Moyer-Gusé et al., 2019). Exposure to narratives featuring outgroup members thus seems to have potential for prejudice reduction, especially in contexts where possibilities for direct intergroup contact are limited (Wojcieszak et al., 2020). Furthermore, several studies have demonstrated the efficacy of written narratives, films, and television series in combating stigma against marginalized groups (Igartua & Frutos, 2017; Igartua et al., 2019; Müller, 2009). A recent meta-analysis also identified 12 studies in the past decade that have used entertainment interventions to reduce prejudice, and found that the meta-analytic effect is quite strong (Paluck et al., 2021). This field of research, termed 'narrative persuasion,' explores how we can use stories to evoke shifts in beliefs, attitudes, and behaviours (Green & Brock, 2000; Moyer-Gusé, 2008).

[3] Defined as feeling sympathy with another person's distress, usually involving a desire to help or comfort them (APA Dictionary of Psychology, 2024c)

[4] Defined as understanding a person from their frame of reference, or vicariously experiencing that person's feelings, perceptions, and thoughts (APA Dictionary of Psychology, 2024d)

Positive media depictions of disadvantaged groups have proved effective in reducing prejudice towards said groups in several studies (Igartua & Frutos, 2017; Igartua et al., 2019; Müller, 2009; Murrar & Brauer, 2018). In some cases, this has even been more effective in reducing prejudice than other established methods of prejudice reduction, such as imagined contact (Murrar & Brauer, 2018). One study found that viewing a 4-minute music-video that portrayed Arabs/Muslims as relatable and likeable resulted in a larger reduction in prejudice than an imagined contact exercise (Murrar & Brauer, 2018). In this study, the imagined contact exercise consisted of the participants being prompted to imagine a positive interaction with a Muslim, and to write about that interaction. This was also compared to a 'group malleability article', consisting of a short article that discussed scientific research regarding ethnic and religious groups changing over time (Murrar & Brauer, 2018). Additionally, the same study found that increased identification with the target group members was associated with greater prejudice reduction (Murrar & Brauer, 2018). Moreover, a recent review established that there is substantial empirical support for the effect of narratives on reducing stigma through text as well as video (Zhuang & Guidry, 2022). It appears that narratives constructed through a first-person point of view are also particularly effective (Chen et al., 2017; Zhuang & Guidry, 2022). This is thought to be due to stories using the first-person point of view being seen as more personal, which in turn increases perspective-taking and identification (Chen et al., 2017; Zhuang & Guidry, 2022).

When it comes to theatre as a narrative medium for reducing stigma, the existing body of research is limited. However, considering the evidence that exists for literary narratives as well as video narratives, it has been suggested that theatre could have similar effects. A study by Rathje et al. (2021) found that, after watching a theatre show people reported greater empathy for the groups depicted in the shows, held opinions that were more consistent with socio-political issues highlighted in the shows, and donated more money to charities related to the shows. Furthermore, in another study where middle schoolers went on field trips to view theatre shows, the participants displayed higher levels of social tolerance as well as perspective-taking as a result (Greene et al., 2018). Thus, based on this evidence, it appears that narratives delivered in a theatre setting could reduce stigma and prejudice.

Using Acceptance and Commitment Therapy to increase flexibility in relation to stigma

Acceptance and Commitment Therapy (ACT) is a mindfulness-based behavioural therapy, which often employs experiential exercises and value-guided behavioural interventions (Hayes et al., 1999; Hayes et al., 2011). The ACT approach to psychological intervention can be defined in terms of six psychological processes that revolve around one

core concept: psychological flexibility. These processes are (1) being present, (2) values, (3) acceptance, (4) defusion (i.e., shifting attention away from the *content* of thoughts to the *process* of thinking), (5) noticing self and (6) committed action. The core concept of psychological flexibility, in turn, can be defined as consciously engaging with the current moment without avoidance, accepting it as it is, and adapting one's behaviour to align with one's values (Hayes et al., 2011). The six processes centered around psychological flexibility form the so-called 'Hexaflex', the key model of ACT used when applied in a therapeutic context (Hayes et al., 2011).

ACT is mainly used as a form of therapy addressing mental health issues such as depression, anxiety, or substance abuse (Hayes, et al., 2011). Nevertheless, many of the psychological processes highlighted within ACT could also be applied to behavioural change in a broader sense (Matsuda et al., 2020). In the context of trying to reduce prejudice and discrimination, the processes of identifying and highlighting one's core values, and committing to making one's actions align with these values, are highly relevant. As previously mentioned, the 'self-regulation of prejudice model' involves making people aware of discrepancies between their prejudices and core values (Monteith et al., 2016).

Psychological flexibility encompasses the ability to actively acknowledge and accept one's internal experiences in the present moment while flexibly engaging or disengaging in behaviours aligned with personal values (Hayes et al. (2011). Psychological inflexibility, on the other hand, is the inability to view internal reactions (such as thoughts and emotions) as transient processes (Hayes et al., 2011; Levin et al., 2014). In other words, it is the inability to relate to one's internal experiences as temporary. This promotes an inflexible attitude towards thoughts and feelings, often leading to behaviours that are directly congruent with immediate thoughts and emotions without complex reflection on their truthfulness or consequences. This can also promote the belief that you are defined by your thoughts and emotions, and that these are not changeable. This process is called "fusion" in ACT (Hayes et al., 2011). Within the context of stigma, reducing psychological inflexibility entails a multifaceted process involving: (1) maintaining a flexible awareness of internal experiences, including stigmatizing cognitions; (2) employing cognitive defusion techniques to disentangle from stigmatizing thoughts, thereby perceiving them as transient rather than immutable truths; (3) cultivating a willingness to accept stigmatizing thoughts instead of resorting to ineffective avoidance strategies, such as thought suppression or evading situations; (4) fostering a sense of detachment from self and others to these thoughts and emotions; (5) clarifying valued behavioural patterns in social interactions; and (6) making committed efforts towards

engaging in valued activities despite the presence of stigmatizing thoughts and emotions (Hayes et al., 2002; Lillis & Hayes, 2007).

In the past few years, research has been conducted on how Acceptance and Commitment Training (ACTr) can be used to reduce societal stigma and prejudice (for review, see Matsuda et al., 2020). Investigations into the influence ACTr on societal stigma and prejudice towards diverse demographics have revealed its efficacy in mitigating these (Hayes et al., 2004; Levin et al., 2016; Lillis & Hayes, 2007; Masuda et al., 2007; Wolgast et al., 2024). An empirical investigation employing a randomized control trial demonstrated significant improvements in substance abuse counsellors' attitudes towards their clients over three months following a one-day ACT workshop compared to those undergoing an educational intervention (Hayes et al., 2004). Notably, the impact of ACT on stigma relative to the control condition was mediated by reductions in the "believability" of stigmatizing thoughts, which is a core aspect of psychological flexibility (Hayes et al., 2004). In other words, the ACT workshop had an impact on the extent to which the participants believed in their stigmatizing thoughts, and a reduction in believability was associated with stigma reduction. Similarly, a study by Lillis and Hayes (2007) found that a 75-minute ACT session led to a greater increase in positive behavioural intentions towards racially diverse groups at a one-week follow-up compared to an education control condition. In addition, the effect of the ACT session on positive behavioural intentions was partially mediated by changes in the measure of acceptance and flexibility towards stigma (Lillis & Hayes, 2007). This meant that greater flexibility with stigma was a significant component in the changes in behavioural intentions. Finally, one study found that a 2.5-hour ACT workshop produced significant reductions in stigma towards mental illness irrespective of the level of psychological flexibility of participants, while an education comparison condition only produced improvements in stigma among those participants higher in psychological flexibility (Masuda et al., 2007).

Collectively, these studies underscore the potential utility of focusing on psychological (in)flexibility when trying to reduce stigma and prejudice towards a broad range of groups. Furthermore, decreasing inflexibility with stigma seems particularly meaningful. However, limited research exists on how ACT processes may be integrated into different types of prejudice reduction interventions. Previous research has mainly compared ACT workshops with educational conditions (e.g., Hayes et al., 2004; Lillis & Hayes, 2007; Masuda et al., 2007). Thus, there's an opening within this field to expand into different intervention methods integrated with ACT principles.

Taken together, the evidence from the above-presented research highlights several significant aspects which are incorporated into the present research. Firstly, the previous research on methods of prejudice reduction highlights the key roles that compassion and taking perspectives through conversations about injustice play in this process (Berger et al., 2018; Shih et al., 2009; Vescio et al., 2003). Second, research on the concept of white fragility and counter-reactions to discussions of racism highlights important barriers that exist in the context of prejudice reduction interventions (DiAngelo, 2011; Hill et al, 2021). Third, research on narratives, media, and theatre highlights how stories and artistic communication can evoke perspective-taking, and the potential that narratives have for reducing prejudice and stigma (Igartua & Frutos, 2017; Igartua et al., 2019; Müller, 2009; Murrar & Brauer, 2018; Zhuang & Guidry, 2022). Research on ACT and ACTr interventions highlights how psychological (in)flexibility, core values and committed action are important processes in the context of prejudice reduction and behavioural change (Hayes et al., 2004; Lillis & Hayes, 2007; Masuda et al., 2007; Wolgast et al., 2024). Additionally, this research highlights that reducing psychological inflexibility with stigma may lead to a more open and accepting outlook on one's own stigmas. This could further promote complex reflection on one's own stigmatising thoughts and the behaviours associated with them. Finally, it has been noted that there is a lack of research on interventions integrating media, entertainment and narratives, despite a high interest in these kinds of interventions (Paluck et al., 2021). Furthermore, there does not currently exist any research that has used the format of an ACTr workshop in a theatre setting. Thus, this novel project is breaking new ground within research on how ACTr can be integrated into different contexts.

Current study

This project aimed to create and test the effect of group conversation about injustice using methods from ACTr in a theatre setting to reduce prejudice, psychological inflexibility with stigma and white fragility as well as increase compassion and behavioural intentions towards equality among a sample of Swedish adults. The research questions raised are as follows: (1) Can group conversation about injustice in a theatre setting enhance compassion and behavioural intentions towards equality, and reduce white fragility, psychological inflexibility with stigma and prejudice towards racial minorities in Sweden? (2) Can group conversation about injustice in a theatre setting be more effective than a focus/discussion group for enhancing compassion and behavioural intentions towards equality, and reducing white fragility, psychological inflexibility with stigma and prejudice towards racial minorities in Sweden? Based on findings from previous research, the hypotheses are as follows:

H1: The intervention will increase the scores for compassion in comparison with the control condition.

H2: The intervention will increase the scores for behavioural intentions towards equality in comparison with the control condition.

H3: The intervention will reduce the scores for prejudice towards racial minorities in comparison with the control condition.

H4: The intervention will decrease psychological inflexibility with stigma in comparison with the control condition.

H5: The intervention will reduce the scores for white fragility in comparison with the control condition.

Method

This quantitative study employed a 2x2 mixed between- and within-subjects design. There were two different participant groups: the intervention group (i.e., the conversational theatre group) and the control group (i.e., a focus/discussion group), as well as a pre-and post-measure of the dependent variables (i.e., compassion, psychological inflexibility with stigma, behavioural intentions towards equality, white fragility and prejudiced attitudes). Furthermore, within each respective group, the participants were split into 10 smaller groups of around 10-15 people (5 intervention and 5 control). This group size has been used previously (e.g., Masuda et al., 2007) and was therefore deemed appropriate for the context of this intervention. The study took place in Malmö, Sweden, in collaboration with the Inter Art Centre at Lund University, an interdisciplinary research centre.

Participants

Participants were recruited from the public through an ad posted on social media, e-mails that were sent through the Lund University network as well as posters in physical spaces. In accordance with the ethical application (Dnr 2018/567), participants had to be above the age of 16. In addition, participants needed to have a good understanding of Swedish since both the intervention and the control condition were conducted in Swedish. In total 85 participants took part in this study. Demographic data for the participants is presented in Table 1.

Table 1.

Demographic data for participants.

	Intervention		Control		Total sample	
Variable	n	%	n	%	n	%
Gender						
Woman	25	62.5	26	57.8	51	60.0
Man	14	35.0	19	42.2	33	38.8
Non-binary	1	2.5	-	-	1	1.2
Age						
18-25	5	12.5	18	40.0	23	27.1
26-35	12	30.0	13	28.9	25	29.4
36-45	10	25.0	4	8.9	14	16.5
46-55	7	17.5	8	17.8	15	17.6
56-65	4	10.0	2	4.4	6	7.1
66-75	2	5.0	-	-	2	2.4
Main occupation						
Working	22	55.0	18	40.0	40	47.1
Student	12	30.0	26	57.8	38	44.7
Unemployed	3	12.5	1	2.2	6	7.1
Homemaker	1	2.5	-	-	1	1.2
Education[a]						
Elementary	2	5.0	2	4.4	4	4.7
High school	6	15.0	6	13.3	12	14.1
University	32	80.0	37	82.2	69	81.2
Ethnicity[b]						
Swedish	31	77.5	35	77.8	66	78.6
Swedish minority[c]	1	2.5	2	4.4	3	3.6
European	7	17.5	7	15.6	14	16.7
North American	1	2.5	1	2.2	2	2.4
South American	-	-	2	4.4	2	2.4
East Asian	-	-	2	4.4	2	2.4
South Asian	1	2.5	1	2.2	2	2.4
South African	1	2.5	1	2.2	2	2.4
Middle eastern	2	5.0	3	6.7	5	6.0

Note. $N = 85$.

[a] Indicates the highest finished level.

[b] Multiple choices were possible.

[c] Jews, Roma people, Sami people, Sweden Finns, and Tornedalians

Measures

Compassion

To measure compassion, a validated Swedish translation of the Compassion Scale (CS; Pommier et al., 2020) was used. In this measure, compassion is operationalized as feeling kindness, general compassion, awareness, and a low degree of indifference to the suffering of others. Examples of items included are "*When others are sad, I try to comfort them*" and *"I notice when people are sad, even when they do not say anything"*. The scale

consists of 16 items and responses are recorded on a 5-point Likert scale (1 = almost never, 5 = almost always). In Pommier et al. (2020), the scale demonstrated good internal consistency ranging between α = .77 and α = .90. In the present study, the scale demonstrated an internal consistency of α = .75.

Psychological inflexibility with stigma

Psychological inflexibility was measured using a validated Swedish translation of the Acceptance and Action Questionnaire – Stigma (AAQ-S; Levin et al., 2014). This is a 20-item measure of psychological flexibility/inflexibility with stigmatising thoughts. The measure consists of two different subscales (psychological flexibility and psychological inflexibility) and includes questions such as "*My biases and prejudices affect how I interact with people from different backgrounds*" and "*I have trouble letting go of my judgments of others*". Responses are recorded on a 7-point Likert scale (1 = never, 7 = always). The flexibility subscale items are reversed, and thus higher scores on the full scale indicate higher inflexibility with stigmatising thoughts. In Levin et al. (2014) the full scale demonstrated a good internal consistency of α = .84. Cronbach's alpha for the psychological inflexibility and psychological flexibility subscales were .85 and .82 respectively. In the present study, the full scale demonstrated an internal consistency of α = .77. Cronbach's alphas for the separate subscales were .83 for psychological inflexibility and .82 for psychological flexibility respectively.

Behavioural intentions

To measure behavioural intentions, the Behaviour Intentions Towards Equality scale (BITE; Hoff & Wolgast, 2024) was used. BITE measures behavioural intentions toward equality and resistance to discrimination (Hoff & Wolgast, 2024). This scale was developed for the Swedish context by Hoff & Wolgast (2024) and the questions were inspired by the items used in Lillis & Hayes (2007). The purpose of the instrument is to measure anti-discriminatory behavioural intentions, focusing on individuals' willingness to react and/or take action against discrimination and the acknowledgement of the consequences of discrimination. It consists of two subscales: Resistance to Discrimination (RD) and Behavioural Intentions (BI). An example of an RD item is *"Swedes and non-ethnic Swedes have different opportunities to succeed because of how society views them."*. An example of a BI item is "*We must protest when we hear simplified descriptions of people of certain ethnicities.*" Responses are recorded on a 5-point Likert Scale (1 = not at all accurate, 5 = accurate). The first version of the scale (used in Ilanius Göransson et al., 2024) consisted of 10 items and the internal consistency was α = .87. Cronbach's alphas for the separate

subscales were RD = .83 and BI = .85. The present study added another five questions and the subscale Engagement in Equality, inspired by items from the Color Blind Racial Attitudes Scale (CoBRAS; Neville et al., 2000). An item example of the EE subscale is *"I do not see any risk that the work against discrimination goes too far, so that ethnic Swedes risk being discriminated against."* In the present study the full scale had an internal consistency of α = .88. Cronbach's alphas for the separate subscales were RD = .77, BI = .89 and EE = .81.

White fragility

To measure white fragility, a measure developed by Hill et al. (2021) was used. This measure consists of two questions; (1) *"When people discuss problems of racism and discrimination and its consequences, what do these discussions make you feel? How often do such discussions make you feel…"* and (2) *"When people discuss so-called white fragility (that white people deny or do not want to see that they have privileges or benefits from being white in society), how do these discussions make you feel? How often do such discussions make you feel…".* These questions each had six response items: (1) confused, (2) drained/exhausted, (3) guilty, (4) angry/annoyed, (5) unsafe and (6) sad. Responses are recorded on a 5-point Likert scale (1 = rarely; 5 = often). These items make up two subscales; Remorse (consisting of the items sad, guilty and angry/annoyed) and Depletion (consisting of the items confused, drained/exhausted and unsafe). In Hill et al. (2021) the measure demonstrated good internal consistency estimates for the Remorse (α = 0.84) and Depletion (α = 0.80) subscales respectively. In the present study, four additional items were added to reflect positive reactions to discussions of racism and white privilege. This was in accordance with the future research directions presented by Hill et al. (2021). These items were (1) engaged, (2) hopeful, (3) empathetic and (4) compassionate. In the present study, the full scale demonstrated an internal consistency of α = 0.76. Cronbach's alphas for the separate subscales were Remorse = .61, Depletion = .77 and Positivity = .83.

Prejudiced attitudes

Prejudiced attitudes towards racial minorities and immigrants were measured using the Modern Racial Prejudice Scale (MRPS; Akrami et al., 2000). The MRPS was developed specifically for a Scandinavian context, making it appropriate for this project. It consists of two subscales: classic racism and modern racism. For the purposes of this study, only the modern racism subscale was used. This subscale consists of nine items, intending to capture denial of the existence of discrimination, resistance to equality, and negativity towards special treatment of minorities (Akrami et al., 2000). The measure includes questions such as "*Immigrants are becoming too demanding in their requirement for equal rights*" and

"*Discrimination of immigrants is no longer a problem in Sweden*". Responses are recorded on a five-point Likert scale (1= completely disagree, 5 = completely agree). In Akrami et al. (2000) the scale demonstrated an internal consistency of α = .82. In the present study the scale demonstrated an internal consistency of α = .85.

Procedure

When participants registered for the study, they were given the option to select from several available dates to attend. These dates were randomly assigned to be either an intervention session or a control session. Thus, neither the participants nor the study coordinators were in control of which participants ended up in which group. In the end, when enough participants had signed up, 10 sessions (5 intervention and 5 control) were carried out. Once participants had signed up to take part in the study, there was a pre-measure of the dependent variables (i.e., compassion, psychological inflexibility with stigma, behavioural intentions, white fragility and prejudiced attitudes). This was done through participants responding to an online survey designed using Lund University's *SUNET Survey* platform. Participants were sent the survey before they were scheduled to take part in their session. All participants were required to complete this pre-measure to be able to further take part in the study. On the day of the scheduled session, participants in the intervention condition took part in the conversation about injustice condition. As for the control group, they took part in a focus discussion group. A week after the participants had completed their respective sessions, they were sent the same survey as before as a post-measure of the dependent variables.

Intervention group

The intervention group received the intervention on one occasion, which lasted on average for 1,5 hours. Exercises inspired by ACTr were interspersed with narrative and testimonial segments read by an actor on video. The intervention was framed to the participants as "conversational theatre" combined with a psychological study, and explained as a format where people reflect and discuss different topics together. The session leader guided participants through the steps of the ACT Hexaflex (Hayes et al., 2011). The workshop was developed by researchers in the project, and the intervention was tested repeatedly before being used in this study, both with participants knowledgeable about the topics addressed and with laypeople. Testing sessions concluded with feedback rounds to further refine the intervention. A detailed outline of the intervention condition is presented in Table 2 (Appendix).

Control group

In the control condition, participants took part in a focus discussion group. This discussion group was equally framed to the participants as in the intervention group and explained as a format where people reflect and discuss different topics together. The control condition had an average duration of 1,5 hours. During the discussion, topics such as fairness, equality, and how we define these things in our society were discussed. The session started with the discussion leader introducing the topic and explaining the format of the discussion. Then, four different videos of the same actor as in the intervention condition were played. In the videos, the actor presented ideas of fairness, equality and how to build a just society from several different philosophers (e.g., Plato, Aristotle, John Rawls, John Stuart Mill, Immanuel Kant, Karl Marx, and Jean-Jacques Rousseau). In between the videos the session leader instigated discussions regarding the ideas presented in each respective video. The discussion leader asked questions such as *"What are your thoughts on what fairness and equality means?"*, "*Do you believe we have a fair society?"*, *"What is required of us as citizens to create a fair society"* and *"Do you feel like you are treated fairly? Why/Why not?"*. After all the video presentations and participant discussions, the session leader thanked the participants and closed out the session.

Ethical considerations

This study has been approved by the Swedish Ethical Review Board (Dnr 2018/567). Participants received written information about the content and purpose of the study before agreeing to participate and were informed that they could terminate their participation at any time. The participants gave their informed consent before they filled out the first survey and were also given the researchers' contact details to ask any questions that they may have had. No sensitive personal data was collected, and all collected information was treated anonymously and not linked to the participants' names.

During the intervention and control condition sessions, the session leader ensured that the conversation remained respectful and in good faith. After filling out the second survey, the participants were given a brief description of the two different conditions, so that they were able to identify which one they took part in. Finally, all participants were offered to be sent the results of the study, and invited to come to a follow-up session where the study results will be presented in detail.

Data analysis

The data was deemed to be normally distributed from an inspection of histograms. To answer the research questions and test the hypotheses posed, repeated measures analyses of

variance (ANOVAs) were performed with time of assessment (pre-session and post-session) as the within-subject factor and condition (intervention and control) as the between-subject factor. The data was examined to confirm that the assumptions for the analysis were met. One outlier was identified in the MRPS measure and was trimmed from the data. Partial eta-squared values were computed as measures of effect size. Significant effects were followed up by post hoc pairwise comparisons with Bonferroni adjustments for multiple comparisons. All analyses were performed using IBM SPSS Statistics version 29.

Results

Descriptive data

Table 3 presents correlations and descriptive statistics for the dependent variables. Out of the 85 participants, one outlier was trimmed and 11 failed to complete the post-measure. Thus 73 participants (86% of the sample) had complete data to be analysed. Of these 35 were in the intervention group and 38 were in the control group.

Table 3.

Correlations and descriptive statistics for the dependent variables.

Variable	*n*	*M*	*SD*	1	2	3	4	5
1. AAQ-S	84	3.06	.58	—				
2. CS	84	4.29	.41	−.33**	—			
3. MRPS	84	1.72	.68	.11	−.16	—		
4. BITE	84	4.05	.62	−.17	.22*	−.80**	—	
5. WF	84	2.69	.54	.03	−.16	.23*	−.29**	—

Note. *p < 0.05, **p < 0.01. Correlations are Pearson's r coefficients. AAQ-S = Psychological inflexibility with stigma. CS = Compassion scale. MRPS = Modern racial prejudice scale. BITE = Behavioural intentions towards equality. WF = White fragility.

Table 4 presents descriptive statistics for the dependent variables for each group at the two time points. The two groups did not differ significantly from each other on any dependent variable score at baseline.

Table 4.

Descriptive statistics for the dependent variables across the two groups.

	Intervention group			
	Before (n = 39)		After (n = 35)	
Variable	*M*	*SE*	*M*	*SE*
AAQ-S	3.20	.57	2.98	.72
Flexibility [a]	3.36	1.06	3.35	1.18
Inflexibility	3.07	.87	2.68	.97
CS	4.26	.41	4.23	.66
MRPS	1.65	.75	1.79	.59
BITE	4.06	.71	4.08	.73
RD	3.87	.77	3.89	.76
BI	4.39	.84	4.38	.88
EE	3.93	1.05	3.98	.99
WF	2.64	.58	2.80	.52
Remorse	2.79	.92	3.02	1.11
Depletion	2.10	.85	2.47	.94
Positivity [a]	2.94	.86	2.88	.84
	Control group			
	Before (n = 45)		After (n = 38)	
Variable	*M*	*SE*	*M*	*SE*
AAQ-S	2.95	.57	2.86	.62
Flexibility [a]	3.03	.70	2.88	.68
Inflexibility	2.88	.71	2.84	.69
CS	4.26	.41	4.22	.52
MRPS	1.80	.62	1.96	.57
BITE	4.03	.53	3.95	.51
RD	3.78	.76	3.59	.75
BI	4.36	.56	4.35	.57
EE	3.96	.63	3.91	.69
WF	2.73	.50	2.73	.50
Remorse	2.58	.66	2.49	.81
Depletion	2.33	.80	2.21	.71
Positivity [a]	3.15	.78	3.31	.85

Note. AAQ-S = Psychological inflexibility with stigma. CS = Compassion scale. MRPS = Modern racial prejudice scale. BITE = Behavioural intentions towards equality. BITE subscales: RD = Resistance to discrimination, BI = Behavioural intentions, EE = Engagement for equality. WF = White fragility.

[a] Reversed scale.

Hypothesis 1

The first hypothesis posed was that the intervention would increase the participants' scores for compassion in comparison with the control condition. Results showed that there was no significant main effect of time, $F(1,71) = 0.344$, $p = .560$, $\eta_p^2 = .005$. There was also no significant effect of condition, $F(1,71) = 0.117$, $p = .734$, $\eta_p^2 = .002$, and no significant

interaction effect, $F(1,71) = 0.413$, $p = .522$, $\eta_p^2 = .006$. This indicates that neither the intervention nor the control condition had any effect on participants' compassion.

Hypothesis 2

The second hypothesis was that the intervention would increase the scores for behavioural intentions towards equality in comparison with the control condition. Results showed that there was no significant main effect of time, $F(1,71) = 0.124$, $p = .726$, $\eta_p^2 = .002$. There was also no significant effect of condition, $F(1,71) = 0.521$, $p = .473$, $\eta_p^2 = .007$, and no significant interaction effect, $F(1,71) = 0.041$, $p = .841$, $\eta_p^2 = .001$. Further analysis of the BITE subscales did not reveal any effects. This indicates that neither the intervention nor the control condition had any effect on participants' behavioural intentions towards equality.

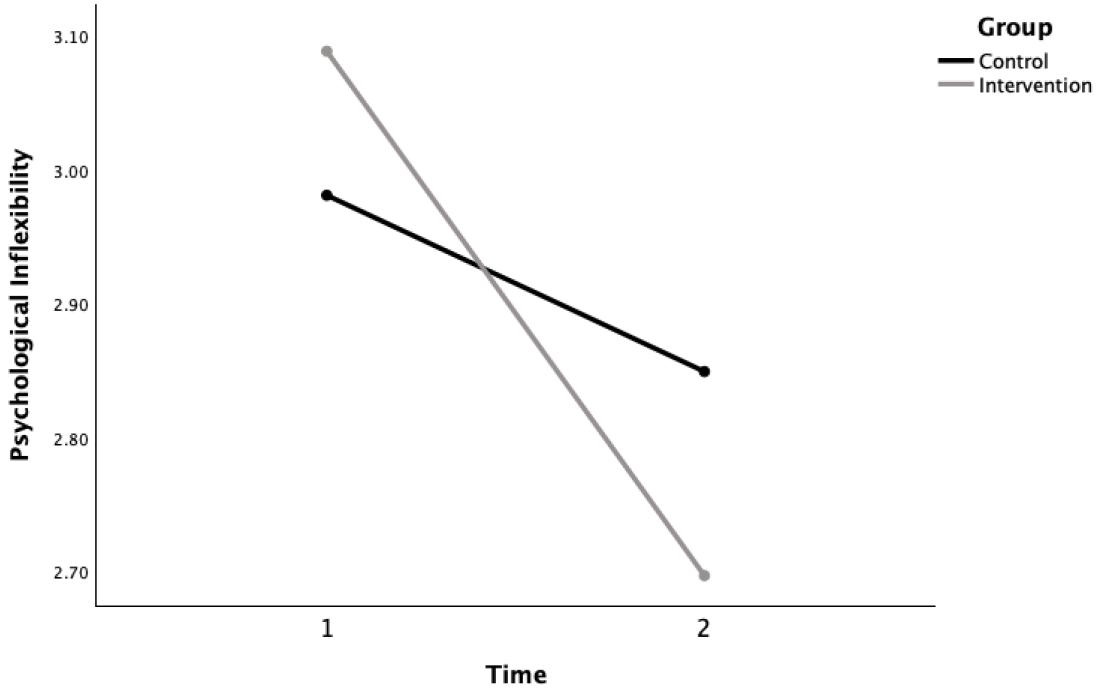
Hypothesis 3

The third hypothesis posed was that the intervention would reduce the scores for prejudice towards racial minorities in comparison with the control condition. Results showed that there was a strong significant main effect of time, $F(1,71) = 7.981$, $p = .006$, $\eta_p^2 = .101$, indicating that levels of prejudice were impacted in both groups as a result of both conditions. Post-hoc pairwise comparisons revealed that prejudice increased in both groups at Time 2 compared to Time 1. There was no significant main effect of condition, $F(1,71) = 1.112$, $p = .295$, $\eta_p^2 = .015$. There was also no significant interaction effect of condition x time, $F(1,71) = 0.002$, $p = .966$, $\eta_p^2 = .002$, indicating no difference in the effect on prejudice between the intervention and control conditions. Thus, prejudice did not differ in the groups at baseline, but increased for participants in both groups.

Hypothesis 4

The fourth hypothesis was that the intervention would decrease psychological inflexibility with stigma in comparison with the control condition. The results showed that there was a strong significant main effect of time, $F(1,71) = 8.698$, $p = .004$, $\eta_p^2 = .109$, indicating that both conditions had an effect on psychological inflexibility. Post-hoc pairwise comparisons revealed a significant decrease in psychological inflexibility at Time 2 compared to Time 1 for both groups. There was no main effect of condition, $F(1,71) = 1.312$, $p = .256$, $\eta_p^2 = .018$, and no significant interaction effect of condition x time, $F(1,71) = 0.885$, $p = .350$, $\eta_p^2 = .012$. However, further analysis of the subscales revealed that there was a moderate significant interaction effect of time x condition for the *inflexibility* subscale, $F(1,71) = 4.574$, $p = .036$, $\eta_p^2 = .060$, but no significant effects for the *flexibility* subscale, $F(1,71) = 0.467$, $p = .496$, $\eta_p^2 = .006$. This indicates that the intervention had a different effect on psychological inflexibility compared to the control condition. The interaction effect is illustrated in Figure 1.

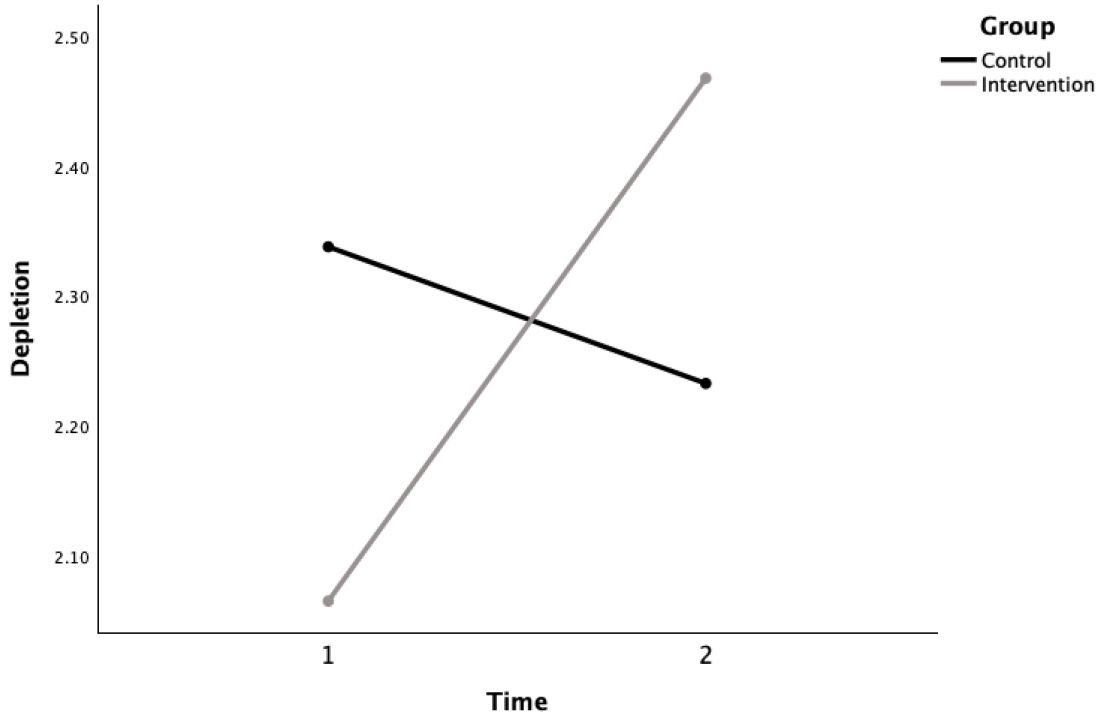
Figure 1. *Mean scores on the psychological inflexibility subscale at the two time points.*



Hypothesis 5

The fifth and final hypothesis was that the intervention would reduce the scores for white fragility in comparison with the control condition. The results showed that there was a small to moderate main effect of Time, $F(1,71) = 3.814$, $p = .049$, $\eta_p^2 = .051$, but no significant main effect of condition, $F(1,71) = 0.069$, $p = .794$, $\eta_p^2 = .001$. However, the condition x time interaction was significant, $F(1,71) = 4.067$, $p = .048$, $\eta_p^2 = .054$, indicating that there was a small to moderate effect of time dependent on condition. Post-hoc pairwise comparisons revealed that there was no significant difference between the groups at baseline and no significant change in the control group over time. However, in the intervention group, white fragility increased as a result of the intervention. Further analysis of the white fragility subscales revealed that there was a moderate significant interaction effect of time x condition for the *depletion* subscale, $F(1,71) = 7.696$, $p = .007$, $\eta_p^2 = .098$, but no significant effects for *remorse*, $F(1,71) = 2.879$, $p = .094$, $\eta_p^2 = .038$, nor *positivity*, $F(1,71) = 1.299$, $p = .258$, $\eta_p^2 = .018$. This indicates that depletion (but not remorse or positivity) increased as a result of the intervention but not the control condition. The interaction effect is illustrated in Figure 2.

Figure 2. *Mean scores for the depletion subscale at the two time points.*



Discussion

The aim of this study was to test the effects of conversation about injustice using methods from ACTr in a theatre setting on compassion, psychological inflexibility with stigma, behavioural intentions towards equality, prejudiced attitudes, and white fragility. The hypotheses were that the intervention (but not the control condition) would positively affect compassion and behavioural intentions and negatively affect prejudice, psychological inflexibility and white fragility. In other words, compassion and behavioural intentions were expected to increase. In contrast, prejudice, psychological inflexibility and white fragility were expected to decrease.

Compassion and behavioural intentions towards equality

Findings did not support the hypotheses regarding compassion nor behavioural intentions, as no significant effects were observed in either group. The lack of effects on compassion and behavioural intentions could possibly be explained by the content of the intervention, which was overall more geared towards reducing prejudice, exposing participants to their own and others' privileges and stereotypes, and illuminating issues of inequality. Perhaps the operationalisation of perspective-taking within this study, such as listening to a testimonial narrative as well as listening to the perspectives of other people in the group during the discussion, was not enough to evoke greater compassion. Conversation

about injustice can be done at different levels. This includes everything from imagining putting oneself in another's shoes to undergoing the experiences of someone else (Paluck et al., 2021). Usually, methods where people are exposed to the experience of someone else are more effective compared to imaginative methods (Paluck et al., 2021). Based on previous findings from research on narrative persuasion, it was expected that the narrative presented in the intervention condition would evoke perspective-taking and consequently increase compassion (Igartua & Frutos, 2017; Igartua et al., 2019; Moyer-Gusé et al., 2019; Murrar & Brauer, 2018). These effects were not observed. However, no actual measure of perspective-taking was included in this study. This was not incorporated since the measure of compassion was deemed to be more relevant. Moreover, earlier research showing that perspective-taking exercises can lead to increased compassion was relied upon (Condon & DeSteno, 2017). Nevertheless, this means that it is not possible to state whether the intervention increased perspective-taking or not, and should perhaps have been included for that purpose. It is possible that the narrative did evoke perspective-taking, despite not affecting compassion, but this is purely speculative. Furthermore, in previous studies where narratives have been found to increase perspective-taking and compassion, the participants were only exposed to the narrative in question (Igartua & Frutos, 2017; Igartua et al., 2019; Murrar & Brauer, 2018). This is rather different from the intervention scenario in the present study, where several sections were interactive and did not consist of a narrative. Thus, previous studies on narrative and compassion can perhaps not be compared to the present study. Moreover, some sections of the intervention condition, such as the hand-raising exercises and discussions, were arguably not mainly targeting compassion or behavioural intentions towards equality. For this intervention to have an impact on compassion, perhaps the narrative needed to be more in focus, or the exercises needed to be more actively targeting compassion.

Another possible explanation for the results on compassion and behavioural intentions is that these two variables were positively skewed in the sample to begin with. Participants in both groups had very high scores at baseline for both compassion and behavioural intentions towards equality. Hence, there could possibly be a ceiling effect present, which could be why no significant changes were observed. This could have arisen due to the recruitment method (i.e., convenience sampling) not being optimal for obtaining a diverse sample of participants representative of the population. Moreover, it is likely that people who are more invested in questions of equality and discrimination were more likely to sign up for the study. Consequently, this could be one reason that certain variables turned out to be skewed, and consequently why no changes were observed.

Prejudice

Unexpectedly, prejudice increased in both groups. This result was opposite to the hypothesised effect. This could be explained as a backlash effect, indicating that participants had a negative reaction to the experience and that their scores on prejudice increased as a result. Discussing group differences and racial discrimination can be positive in terms of improving attitudes, but might also run the risk of reinforcing stereotypes, and 'backfire' by increasing and/or reinforcing existing prejudices (Paluck, 2006). Moreover, previous research has found that those who are in a majority group (in this case people who are white) may associate equality and multiculturalism with exclusion (Plaut et al., 2011). Seeing as the majority of participants in this study (78.6%) were white Swedes, it is possible that a similar effect was present here, which could explain the increase in prejudice. On the other hand, increased scores for prejudice do not have to indicate that the participants' prejudices changed or became stronger over time. This result could simply reflect an increased *awareness* among the participants of their prejudices. People may have scored higher on the measure of prejudice because they became more aware of their prejudices, through the process of the intervention. Moreover, it is important to note that becoming more aware of one's prejudices does not always equate to the content of these prejudices changing. Thus, increased scores on the prejudice questionnaire cannot always be said to indicate that the participants in fact became *more* prejudiced. In this sense, questionnaires measuring prejudice are limited and perhaps insufficient, since they cannot reflect how people relate to, or reason with, their prejudiced thoughts. Increased scores could therefore be misleading in this type of context.

Observing increased awareness of one's prejudice is in accordance with ACT theory (Hayes et al., 2011). Increasing awareness of prejudice, and admitting to having prejudiced thoughts, is precisely the process that ACT and ACTr intend to initiate (Levin et al., 2016; Lilis & Hayes, 2007; Matsuda et al., 2020). Through ACT and ACTr, people should become more aware of their thoughts, feelings and prejudices (Hayes et al., 2011; Lilis & Hayes, 2007; Matsuda et al., 2020). The key point, however, is that one should be able to *notice* negative thoughts and feelings, but not *act* according to them (Hayes et al., 2011; Lilis & Hayes, 2007). Moreover, ACT theory suggests that we need to get in touch with thoughts and feelings that we avoid, but is not mainly looking to change the content of thoughts or feelings (Hayes et al., 2011). Rather, what is important is awareness and understanding that the presence of negative thoughts and/or prejudices does not need to lead to action (Hayes et al., 2011). Taking this into account, the results on prejudice observed here could still be in line with what ACTr workshops want to achieve, which is increased awareness and acceptance of

prejudiced thoughts. Additionally, the fact that psychological inflexibility with stigma also seemed to decrease in both groups would further indicate that the participants had a higher acceptance of their stigmatising thoughts as a result of the sessions. However, the results from the present study also contradict previous findings, where prejudice and stigma were found to decrease following an ACTr workshop (Lillis & Hayes, 2007; Wolgast et al., 2024). It is possible that the single session implemented in this intervention was not sufficient, and several sessions might have been necessary for an observable effect on the intervention group. Wolgast et al. (2024), for example, implemented three separate sessions. Lillis and Hayes (2007), however, also only implemented one session. Furthermore, it is important to note that prejudice increased in both groups. The control condition did not use methods from ACTr and was not focused on discussions of racial inequality, but rather "justice" as a broader concept. This result could be explained by the fact that the control condition was equally as effective an 'intervention' as the intervention group. Nevertheless, it should be considered that the effects might not be caused by the study conditions. External circumstances and events, or variables not accounted for within this study, could also have affected the participants and be another explanation for the results.

Psychological inflexibility with stigma

For psychological inflexibility, the findings supported the hypothesis that the intervention would lead to decreased psychological inflexibility with stigma. However, the hypothesis that the control condition would not have the same effect on psychological inflexibility was not fully supported, as the results indicate that the control condition also had a negative effect. This was not expected since the control condition did not include any methods from ACTr, which are the methods that have previously been found to impact psychological (in)flexibility (Hayes et al., 2004; Lillis & Hayes, 2007). However, currently there is not any existing research investigating the effect of other types of interventions (i.e., not ACTr-based interventions) for increasing psychological flexibility with stigma. Thus, the findings from the present study are not necessarily opposing any previous findings. This evidence merely indicates that the design of the control condition in the present study also seemed to have a negative effect on psychological inflexibility with stigma. In other words, it also seemingly decreased the participants' psychological inflexibility. Perhaps this could be the control condition, similarly to the intervention condition, offering participants an open forum for discussions of stigmatised topics. This could have evoked similar thought processes as the intervention condition, even though the control condition did not engage participants to the same extent in terms of confronting their own stigmas.

When examining the subscales the results showed that the intervention had a significantly stronger effect on inflexibility than the control condition. Coupled with the observed increase in prejudice, the observed decrease in psychological inflexibility with stigma also supports the interpretation that the participants became more aware of their prejudices. It could also indicate that, while they became more aware of their prejudices, these were not increased or reinforced. Rather, it seems as though participants became more accepting of their negative thoughts (i.e. willing to have the thoughts and emotions without thinking that they are true or having an urge to react to them). As previously mentioned, psychological inflexibility refers to patterns of fused thoughts and emotions where actions are rigidly guided by thoughts, feelings, and urges, rather than personal values (Hayes et al., 2011; Levin et al., 2014). In other words, if one is highly psychologically inflexible, one may tend to act based directly on how one thinks or feels in the moment rather than what would be most effective or meaningful (Hayes et al., 2011; Levin et al., 2014). Observing a decrease in psychological inflexibility would therefore indicate that people might have become more able to reflect on the stigmatizing thoughts that they were confronted with through the intervention condition. Consequently, they may have also become less likely to act on these thoughts and feelings. However, to assert whether this was the case, follow-up testing would have been needed, which was unfortunately out of the scope of this project.

Since a change was observed in both groups, it cannot be ruled out that there were other reasons behind the observed effect nor that it could be a placebo effect. Nevertheless, the significant interaction effect observed for the inflexibility subscale does indicate that the intervention condition was more effective at decreasing psychological inflexibility with stigma compared to the control condition. As previously mentioned, this is in line with previous research where ACTr has been used to influence psychological (in)flexibility (Hayes et al., 2004; Lillis & Hayes, 2007).

White fragility

A final noteworthy finding is that white fragility seemingly increased in the intervention group but not in the control group. This result is also opposite to what was hypothesised since the intervention was assumed to decrease white fragility. However, further investigation revealed that it was specifically the depletion subscale of fragility that showed a significant increase after the intervention. This indicates that the participants in the intervention condition experienced feeling drained, confused and/or unsafe in relation to discussions of racism after the intervention session. The intervention included several exercises where participants had to think and talk about difficult topics such as privileges,

stereotypes, and prejudice. Furthermore, the participants listened to a rather graphic testimonial narrative where experiences of racism were described. Considering this, it is understandable that the intervention condition increased the depletion dimension of white fragility, since these are quite touching topics.

Increased scores for white fragility coupled with increased prejudice could further paint the picture of the intervention potentially having a backfiring effect. However, another interpretation of the results is that the intervention increased the participants' awareness of their reactions and highlighted their part in perpetuating racialized inequality. In other words, they might have become more aware of their problematic reactions and behaviours. As previously mentioned, through ACT, people are meant to become aware of their thoughts and feelings (Hayes et al., 2011). Taking this into account, it might be reasonable that people increase in depletion feelings due to their increased awareness of their biases. This would be in line with the self-regulation of prejudice model (Monteith et al., 2016). It is also supported by previous research that has found that people who value egalitarianism and desire to be unbiased tend to experience negative emotions when faced with the "reality" of their prejudiced attitudes (Fehr & Sassenberg, 2010). Consequently, perhaps it is not possible to increase awareness without also increasing fragility in some way. Some negative emotions may always be present in this process. Moreover, this may be especially true for people whose personal values and ideas of themselves include caring about equality (Fehr & Sassenberg, 2010; Monteith et al., 2016). Since many of the participants in the present study scored high on the compassion and behavioural intentions toward equality scales, they may have been likely to experience similar negative emotions a consequence the intervention. Possibly, including a measure of guilt in this study could have been useful to examine these reactions. If it had increased, it would have supported the above interpretation. While the measure of white fragility used in this study did include "guilty" as one of the response items, this is arguably not sufficient to capture reactions of guilt.

Taken together, the results of the present study paint a complex picture. The intervention condition seemingly affected the participants in a significant way, where their scores for prejudice and white fragility increased and psychological inflexibility decreased. At the same time the control condition seemingly also had an effect, where prejudice was observed to increase and psychological flexibility decreased, while white fragility was not impacted. Moreover, the effects on psychological inflexibility were stronger in the intervention group than in the control group. These results contradict some previous findings regarding the impact of ACTr on prejudice and stigma (Lillis & Hayes, 2007; Wolgast et al.,

2024), but also corroborate some findings regarding the emotional reactions people may have to discussions of discrimination (Fehr & Sassenberg, 2010; Monteith et al., 2016).

Limitations

This project is not without its limitations. The nature of the questions asked in the surveys, as well as some of the questions asked during the discussions in the intervention condition, have a high risk of social desirability in their responses. In other words, it is possible that participants will not answer honestly, but rather in line with what they think they should respond, to appear as a good person. Additionally, the recruitment methods used for this study were not optimal for obtaining a diverse and representative sample of participants. Furthermore, the sample size in this study ($N = 73$), would not be considered large. Both things affect the generalisability of the results.

Another limitation is that it was not possible to control for all the possible confounding variables in the statistical analysis, such as the fact that each group had a different discussion. It is therefore not possible to say that there are indeed causal effects of the intervention or the control condition on the variables where effects were observed. Furthermore, due to the nature of data collection (i.e., online surveys e-mailed to the participants to fill out in their own time) it was not possible to ensure that all participants filled out the surveys within the same time interval. Some participants filled out the first survey a few days before their session, while others filled out the survey right before their session. Additionally, some participants filled out the second survey one week after their session (right as it was sent to them) while others took as long as three weeks to respond to the second survey. Consequently, some participants had longer times between their first and second time filling out the survey, while other participants had a shorter time. This leaves room for error in the interpretation of the results, as we cannot be sure whether these time interval differences affected the results. It is possible that participants who had a longer time between their session and filling out the second survey did not recall as well what transpired during their session and therefore had “weaker” responses than those who filled out the second survey after one week. Certain immediate effects that the sessions may have had on participants could thus have been missed.

As previously mentioned, only having one session meant that there was no time to go into how the participants could deal with the thoughts and feelings brought on by the intervention session. Having several instances where the participants could continue to work through the processes initiated might have been necessary for achieving changes in their attitudes. Furthermore, it should be considered that the scales used (e.g., the Modern Racial

Prejudice Scale and the White Fragility Scale) are not suitable for interventions aimed at participants discovering their own negative thoughts, emotions, and reactions. They could be misleading since they might show elevated levels that do not necessarily reflect the intervention having a negative impact. Furthermore, these types of measures do not reflect how people relate to their thoughts and reactions. Through previous research, we know that when people allow themselves to be exposed to negative thoughts and feelings, they can also start to re-evaluate/think differently about them (Levin et al., 2016; Lilis & Hayes, 2007; Matsuda et al., 2020; Wolgast et al., 2024). However, this was not developed enough in this format due to its short duration. This should be considered in future research, where perhaps a qualitative measure could be included to assess the participants' subjective experiences.

Finally, there are certain limitations to the design of this study. It is important to note that this study employed a quasi-experimental design, meaning that the participants were not completely randomized to their conditions. Without random assignment, there is a risk of selection bias where certain types of participants self-select into different groups based on characteristics that may influence the outcome. This also means that groups may not be equivalent at the outset, which can introduce bias and confounding variables. While the participants did not know which group (intervention or control) they were selecting when they signed up, it is still a significant limitation of this study. Moreover, since there was no idle control group (i.e., a control group who were not subjected to any kind of session) we cannot say with certainty that it was in fact the study sessions that impacted the participants and created changes in the dependent variables. It is also not possible to know what drives the effects, nor to rule out the possibilities of repeated testing effects, history effects (i.e., something happening during the time of the intervention), or confounding variables not controlled for.

Future research

Several hypotheses would have been interesting to explore in this study but were ultimately out of the scope of this project. It would have been interesting to examine possible mediators and/or moderators on the effects of the intervention. For example, it would have been interesting to look at whether ethnicity could be a moderating variable in the effect of the intervention and/or control condition on the dependent variables. This is important to consider since prejudice towards immigrants as well as displays of white fragility likely manifest very differently for people of different ethnicities (e.g., white Swedes compared to non-white Swedes with an immigrant background). White fragility, in particular, is a construct that is assumed to manifest mainly for white individuals, and not for people of

colour (DiAngelo, 2011). However, since the measure of white fragility used in this study included the subscales of remorse, depletion and positivity, it can still be considered applicable to people who are not white. This measure indicates different ways that people respond to discussions of racism, inequality and white privilege, which is also relevant for people of other ethnicities. Non-white people may tend to feel less remorse (i.e., feeling sad or guilty) but more depletion (i.e., feeling drained, exhausted, or unsafe) in the contexts of these discussions. Future research should investigate these intricacies and examine how ethnicity may moderate the effects of an intervention such as the one used in this study.

Another exploratory hypothesis for future research is that higher levels of white fragility and prejudice at baseline could moderate the effect of the intervention on prejudice reduction and compassion enhancement. For example, highly prejudiced participants could be likely to have a negative backlash reaction to the intervention and thus become more prejudiced as a result. Similarly, the intervention may be less effective among people with high white fragility and resistance to discussions of injustice.

Since previous research has stressed the importance of developing prejudice reduction methods with lasting effects (Paluck et al., 2021), future research should also include extended follow-up measurements. This study has not examined long-lasting effects since it was out of the scope of this project. However, future longitudinal studies with extended follow-up periods could provide insights into the long-term effects of interventions aimed at reducing prejudice. Moreover, longitudinal studies with multiple sessions should be implemented to investigate whether sustained exposure could have different results. Additionally, in future research it could be useful to incorporate exercises targeting self-compassion. Self-compassion refers to being supportive toward oneself when experiencing negative emotions caused by personal mistakes and inadequacies or external life challenges (Neff, 2023). This may be constructive, since the observed effects in this study indicate that the intervention may have caused some negative emotional reactions. Facing one's own prejudices and overcoming them could therefore be a process aided by enacting self-compassion.

Finally, given the results indicating no significant effects on compassion and behavioural intentions, future studies aimed at compassion enhancement should refine the content and focus of the intervention. Additional research should also consider increasing the emphasis on the narrative, and develop interactive exercises specifically targeting compassion and behavioural intentions towards equality. Furthermore, future research could expand to compare the effects of narrative-only interventions, interactive-only interventions and

interactive-narrative combined interventions. This way, it may be possible to determine whether exposure to narratives alone, or in combination with specific activities, could enhance compassion.

Conclusions

This study aimed to assess the impact of conversation about injustice using methods from ACTr within a theatre setting on various psychological and attitudinal variables such as compassion, psychological flexibility with stigma, behavioural intentions towards equality, prejudiced attitudes, and white fragility. The findings revealed mixed results against the initial hypotheses. The study did not find support for the hypotheses related to increased compassion or behavioural intentions. This could be attributed to the nature of the intervention, which predominantly focused on reducing prejudice and increasing inequality awareness. Moreover, ceiling effects from high baseline scores and sampling limitations may have influenced these findings.

Both the intervention and control conditions seemingly led to decreased psychological inflexibility with stigma. However, the intervention condition had a stronger effect. If it was indeed an effect of both conditions, this may challenge assumptions about the necessity of ACT and ACTr methods for this outcome. It could however be the case that the control condition used in the present study, through being an interactive discussion group, also evoked active reflection among participants. In previous ACTr studies, the control condition has usually been an "educational" condition such as an informational lecture (e.g., Hayes et al., 2004; Lillis & Hayes, 2007; Masuda et al., 2007). In this study, the control condition was much more actively engaging. However, since the intervention was observed to have a stronger effect, the results also indicate that ACT and ACTr are especially effective in the context of reducing psychological inflexibility. Nevertheless, these results underscore the potential impact of having open discussions on stigmatised topics, regardless of intervention type.

Levels of prejudice were observed to increase in both the intervention and the control group, which could be a potential backlash effect. However, it could also be an effect of increased awareness of existing biases. Combined with the increase in the white fragility depletion subscale and the decreased scores for psychological inflexibility with stigma in the intervention group, the results of this study can possibly be interpreted as successfully raising the participants' awareness of their prejudice. It may also suggest that both the intervention condition and the control condition enhanced the participants' ability to accept prejudiced thoughts without judgement or action. However, further research needs to be conducted to

clarify and explore these complex effects. Several limitations of the current study (such as the measures used and the limited intervention duration) mean that these interpretations can only be made tentatively at this point.

In summary, this study provides valuable insights into the complexities of conversation-based interventions within a theatre setting. It also offers novel evidence of an original approach to prejudice reduction research, which integrates elements of entertainment (such as theatre and narratives) into psychological research grounded in ACT theory. Further, it contributes to our understanding of how people may be affected by discussing inequality and being faced with topics such as privilege and discrimination. The results underscore the need for continued research to refine methods and deepen our understanding. Future studies should consider moderating factors and assess long-term impacts to develop effective interventions that promote compassion and equality.

References

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Akrami, N., Ekehammar, B., & Araya, T. (2000). Classical and modern racial prejudice: A study of attitudes toward immigrants in Sweden. *European Journal of Social Psychology*, *30*(4), 521–532. https://doi.org/b6jxfw

APA Dictionary of Psychology. (2024a). Retrieved 15 April 2024, from https://dictionary.apa.org/prejudice

APA Dictionary of Psychology. (2024b). Retrieved 15 April 2024, from https://dictionary.apa.org/discrimination

APA Dictionary of Psychology. (2024c). Retrieved 15 April 2024, from xshttps://dictionary.apa.org/compassion

APA Dictionary of Psychology. (2024d). Retrieved 15 April 2024, from https://dictionary.apa.org/empathy

Berger, R., Brenick, A., & Tarrasch, R. (2018). Reducing Israeli-Jewish Pupils' Outgroup Prejudice with a Mindfulness and Compassion-Based Social-Emotional Program. *Mindfulness*, *9*(6), 1768–1779. https://doi.org/10.1007/s12671-018-0919-y

Chen, M., Bell, R. A., & Taylor, L. D. (2017). Persuasive effects of point of view, protagonist competence, and similarity in a health narrative about type 2 diabetes. Journal of Health Communication, 22(8), 702–712. https://doi.org/10.1080/10810730.2017.1341568

Condon, P., & DeSteno, D. (2017). Enhancing compassion: Social psychological perspectives. In E. M. Seppälä, E. Simon-Thomas, S. L. Brown, M. C. Worline, C. D. Cameron, & J. R. Doty (Eds.), *The Oxford handbook of compassion science.* (pp. 287–298). Oxford University Press.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*(6), 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003

DiAngelo, R. (2011). White fragility. *International Journal of Critical Pedagogy, 3*(3), 54–70.

Diskrimineringsombudsmannen. (2022, March 24). *Statistik 2015–2021: Statistik över anmälningar som inkom till Diskrimineringsombudsmannen 2015–2021.* (Report no. 2022:2) https://www.do.se/kunskap-stod-och-vagledning/publikationer-om-diskriminering/2023/statistik-over-anmalningar-till-do-2015-2022

Dover, T. L., Hunger, J. M., & Major, B. (2020). Health Consequences of Prejudice and Discrimination. In *The Wiley Encyclopedia of Health Psychology* (pp. 231–238). John Wiley & Sons, Ltd.

Dovidio, J. F., Pearson, A. R., Gaertner, S. L., & Hodson, G. (2008). On the nature of contemporary prejudice: From subtle bias to severe consequences. In V. M. Esses & R. A. Vernon (Ed.), *Explaining the breakdown of ethnic relations: Why neighbors kill*. Blackwell Publishing. https://doi.org/10.1002/9781444303056.ch3

Dovidio, J. F., Schellhaas, F. M. H., Love, A., & Hewstone, M. (2017). Reducing intergroup bias through intergroup contact: Twenty years of progress and future directions. *Group Processes and Intergroup Relations*, *20*(5), 606–620. https://doi.org/gbvqnb

Fehr, J., & Sassenberg, K. (2010). Willing and able: How internal motivation and failure help to overcome prejudice. *Group Processes & Intergroup Relations*, *13*(2), 167–181. https://doi.org/10.1177/1368430209343116

Gaertner, S., & Dovidio, J. (2005). Understanding and Addressing Contemporary Racism: From Aversive Racism to the Common Ingroup Identity Model. *Journal of Social Issues*, *61*, 615–639. https://doi.org/10.1111/j.1540-4560.2005.00424.x

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Greene, J., Erickson, H., Watson, A., & Beck, M. (2018). The Play's the Thing: Experimentally Examining the Social and Cognitive Effects of School Field Trips to Live Theater Performances. *Educational Researcher*, *47*, 0013189X1876103. https://doi.org/10.3102/0013189X18761034

Golec de Zavala, A., & Cichocka, A. (2012). Collective narcissism and anti-Semitism in Poland. *Group Processes & Intergroup Relations*, *15*(2), 213–229. https://doi.org/10.1177/1368430211420891

Harvard Library. (2024). Retrieved 22 May 2024, from https://library.harvard.edu/confronting-anti-black-racism/scientific-racism

Hayes, S. C., Bissett, R., Roget, N., Padilla, M., Kohlenberg, B. S., Fisher, G., Masuda, A., Pistorello, J., Rye, A. K., Berry, K., & Niccolls, R. (2004). The impact of acceptance and commitment training and multicultural training on the stigmatizing attitudes and professional burnout of substance abuse counselors. *Behavior Therapy*, *35*(4), 821–835. https://doi.org/10.1016/S0005-7894(04)80022-4

Hayes, S. C., & Hofmann, S. G. (2017). The third wave of cognitive behavioral therapy and the rise of process-based care. *World Psychiatry*, *16*(3), 245–246. https://doi.org/10.1002/wps.20442

Hayes, S. C., Niccolls, R., Masuda, A., & Rye, A. K. (2002). Prejudice, terrorism, and behavior therapy. *Cognitive and Behavioral Practice*, *9*(4), 296–301. https://doi.org/10.1016/S1077-7229(02)80023-2

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2011). *Acceptance and Commitment Therapy, Second Edition: The Process and Practice of Mindful Change*. Guilford Press.

Hill, T., Mannheimer, A., & Roos, J. M. (2021). Measuring White Fragility. *Social Science Quarterly*, *102*(4), 1812–1829. https://doi.org/10.1111/ssqu.12985

Hoff, E., & Wolgast, S. N (2024, in progress). *Behavioral Intentions Towards Equality (BITE): Validation of an anti-racism scale that goes beyond just racial attitudes.* [Unpublished manuscript].

Hsieh, W., Faulkner, N., & Wickes, R. (2022). What reduces prejudice in the real world? A meta-analysis of prejudice reduction field experiments. *British Journal of Social Psychology*, *61*(3), 689–710. https://doi.org/10.1111/bjso.12509

Igartua, J.-J., & Frutos, F. j. (2017). Enhancing attitudes toward stigmatized groups with movies: Mediating and moderating processes of narrative persuasion. *International Journal of Communication*, *11*(1), 158–177.

Igartua, J.-J., Wojcieszak, M., & Kim, N. (2019). How the interplay of imagined contact and first-person narratives improves attitudes toward stigmatized immigrants: A conditional process model. *European Journal of Social Psychology*, *49*(2), 385–397. https://doi.org/10.1002/ejsp.2509

Ilanius Göransson, S., Bladkrona, A., Wolgast, S N, & Hoff E. (2024, in progress) *A Multi-method Approach to Promote Workplace Equality Reduced Prejudice and Strengthened Anti-discriminatory Intentions.* [Unpublished manuscript]. Department of Psychology, Lund University.

Levin, M. E., Luoma, J. B., Vilardaga, R., Lillis, J., Nobles, R., & Hayes, S. C. (2016). Examining the role of psychological inflexibility, perspective taking, and empathic concern in generalized prejudice: Journal of Applied Social Psychology. *Journal of Applied Social Psychology*, *46*(3), 180–191. https://doi.org/10.1111/jasp.12355

Levin, M. E., Luoma, J. B., Lillis, J., Hayes, S. C., & Vilardaga, R., (2014). The Acceptance and Action Questionnaire – Stigma (AAQ-S): Developing a measure of

psychological flexibility with stigmatizing thoughts. *Journal of Contextual Behavior Science,* 3(1), 21–26. doi: 10.1016/j.jcbs.2013.11.003

Lillis, J., & Hayes, S. C. (2007). Applying Acceptance, Mindfulness, and Values to the Reduction of Prejudice: A Pilot Study. *Behavior Modification*, *31*(4), 389–411. https://doi.org/10.1177/0145445506298413

Lowe, M. (2020). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, *111*(6), 1807–1844. https://doi.org/gkcr9m

Masuda, A., Hayes, S. C., Fletcher, L. B., Seignourel, P. J., Bunting, K., Herbst, S. A., Twohig, M. P., & Lillis, J. (2007). Impact of acceptance and commitment therapy versus education on stigma toward people with psychological disorders. *Behaviour Research and Therapy*, *45*(11), 2764–2772. https://doi.org/10.1016/j.brat.2007.05.008

Matsuda, K., Garcia, Y., Catagnus, R., & Brandt, J. A. (2020). Can Behavior Analysis Help Us Understand and Reduce Racism? A review of the Current Literature: Behavior Analysis in Practice. *Behavior Analysis in Practice*, *13*(2), 336–347. https://doi.org/10.1007/s40617-020-00411-4

McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Ed.), *Prejudice, discrimination and racism* (s. 91–126). Academic Press.

Monteith, M. J., Parker, L. R., & Burns, M. D. (2016). The self-regulation of prejudice. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 409–432). Psychology Press.

Moyer-Gusé, E. (2008). Toward a Theory of Entertainment Persuasion: Explaining the Persuasive Effects of Entertainment-Education Messages. *Commun. Theor.* 18 (3), 407–425. doi:10.1111/j.1468-2885.2008.00328.x

Moyer-Gusé, E., Dale, K. R., & Ortiz, M. (2019). Reducing Prejudice Through Narratives. *Journal of Media Psychology*, *31*(4), 185–195. https://doi.org/10.1027/1864-1105/a000249

Müller, F. (2009). Entertaining Anti-Racism. Multicultural Television Drama, Identification and Perceptions of Ethnic Threat: Communications: The European Journal of Communication Research. *Communications: The European Journal of Communication Research*, *34*(3), 239–256. https://doi.org/10.1515/comm.2009.016

Murrar, S., & Brauer, M. (2018). Entertainment-education effectively reduces prejudice. *Group Processes & Intergroup Relations*, *21*(7), 1053–1077. https://doi.org/10.1177/1368430216682350

Pager, D., & Shepherd, H. (2008). The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual Review of Sociology*, *34*, 181–209. https://doi.org/10.1146/annurev.soc.33.040406.131740

Paluck, E. L. (2006). Diversity Training and Intergroup Contact: A Call to Action Research. *Journal of Social Issues*, *62*(3), 577–595. https://doi.org/10.1111/j.1540-4560.2006.00474.x

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*, 339–367. https://doi.org/10.1146/annurev.psych.60.110707.163607

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, *72*, 533–560. https://doi-org.ludwig.lub.lu.se/10.1146/annurev-psych-071620-030619

Park, S. Y. (2012). Mediated Intergroup Contact: Concept Explication, Synthesis, and Application. *Mass Communication and Society*, *15*(1), 136–159. https://doi.org/10.1080/15205436.2011.558804

Pettigrew, T. F., Christ, O., Wagner, U., & Stellmacher, J. (2007). Direct and indirect intergroup contact effects on prejudice: A normative interpretation. *International Journal of Intercultural Relations,* 31(4), 411–425. https://doi.org/10.1016/j.ijintrel.2006.11.003

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, *25*(1), 57–75. https://doi.org/10.1002/ejsp.2420250106

Plaut, V., Garnett, F., Buffardi, L., & Sanchez-Burks, J. (2011). 'What About Me?' Perceptions of Exclusion and Whites' Reactions to Multiculturalism. *Journal of Personality and Social Psychology*, *101*, 337–353. https://doi.org/10.1037/a0022832

Rathje, S., Hackel, L., & Zaki, J. (2021). Attending live theatre improves empathy, changes attitudes, and leads to pro-social behavior. *Journal of Experimental Social Psychology*, *95*, 104138. https://doi.org/10.1016/j.jesp.2021.104138

Shih, M., Wang, E., Bucher, A. T., & Stotzer, R. (2009). Perspective taking: Reducing prejudice towards general outgroups and specific individuals. *Group Processes & Intergroup Relations*, *12*(5), 565–577. https://doi.org/10.1177/1368430209337463

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology*, *100*(6), 1027–1042. https://doi.org/10.1037/a0022308

Vescio, T. K., Sechrist, G. B., & Paolucci, M. P. (2003). Perspective taking and prejudice reduction: The mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology*, *33*(4), 455–472. https://doi.org/drv275

Wolgast, S., Wolgast, M., K. Sandeberg, H. Wasif & Hoff, E. (2024). *Flexibility and Compassion Training for Equality in Organizations: a Randomized Controlled Trial.* [Manuscript submitted for publication].

Wolgast, M., & Wolgast, S. (2021). *Vita privilegier och diskriminering: Processer som vidmakthåller rasifierade ojämlikheter på arbetsmarknaden: Vol. 2021:23*. Länsstyrelsen i Stockholms län.

White, F. A., Borinca, I., Vezzali, L., Reynolds, K. J., Lyshol, B., Verrelli, J. K., S., & Falomir-Pichastor, J. M. (2021). Beyond direct contact: The theoretical and societal relevance of indirect contact for improving intergroup relations. *Journal of Social Issues,* 77, 132–153. https://doi.org/10.1111/ josi.12400

Wojcieszak, M., Kim, N., & Igartua, J.-J. (2020). How to Enhance the Effects of Mediated Intergroup Contact? Evidence from Four Countries. *Mass Communication and Society*, *23*(1), 71–106. https://doi.org/10.1080/15205436.2019.1630444

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology*, 73(1), 73–90. https://doi.org/10.1037/0022-3514.73.1.73

Wölfer, R., Christ, O., Schmid, K., Tausch, N., Buchallik, F. M., Vertovec, S., & Hewstone, M. (2019). Indirect contact predicts direct contact: Longitudinal evidence and the mediating role of intergroup anxiety. *Journal of Personality and Social Psychology*, 116(2), 277–295. https://doi.org/10.1037/ pspi0000146

Zhuang, J., & Guidry, A. (2022). Does Storytelling Reduce Stigma? A Meta-Analytic View of Narrative Persuasion on Stigma Reduction. *BASIC AND APPLIED SOCIAL PSYCHOLOGY*. https://doi.org/10.1080/01973533.2022.2039657

Appendix

Table 2.

Outline of the intervention condition.

Section	Description	Purpose/process
Introduction	A short video where a monologue is performed by an actor. The monologue references Jean Jacques Rousseau's *'Du Contrat Social'* (i.e., the social contract), and highlights that everyone is supposed to have the same rights, opportunities and duties within society.	Introducing the topic of the social contract and setting the scene for the discussions which are about to take place.
Part I	The session leader reads out statements such as "*I usually feel safe*" and "*I am usually listened to and taken seriously*". Participants are asked to raise their hand if they agree with the statement. This is followed by a few discussion questions, regarding what the participants thought of the exercise, what they may have noticed in terms of how different people responded, and their thoughts on what 'privilege' means.	Highlighting privileges. Making participants aware of their own privileges as well as starting a discussion about what it means to have privilege.
Part II	The session leader reads an incomplete sentence and asks the participants to silently finish this sentence with the first word that comes to their mind. The first sentences are innocuous, to familiarize the participants with the exercise, for example "*The sky is…*" and "*Merry Christmas and happy new…*". Then the statements become aimed at activating certain stereotypes, such as for example "*All women wearing a Hijab are…*" and "*All Arabic men are…*". After all sentences are read out, the session leader instigates a discussion asking the participants what words they thought of during the exercise.	Highlighting stereotypes and making the participants aware of their own and others' stereotypical associations.
Part III	A video recording of the actor reciting a personal testimony is played. This testimony highlights racism in the workplace and in everyday life.	Evoking perspective-taking and compassion through narrative.
Part IV	The session leader askes the participants *"What does it mean to be white?",* which is followed by a discussion and subsequently a second hand-raising exercise with new statements such as *"I can be quite certain that if I ask to speak to the manager I will get to speak to someone with the same skin colour/ethnicity as me"* and *"Areas where people with my skin colour/ethnicity live are generally considered 'good areas'".*	Highlighting white privilege and opening up a discussion about 'whiteness'.
Part V	The session leader asks the participants questions such as *"How do you react when*	Highlighting white fragility.

	people bring up that there exists injustice based on race and ethnicity?" and *"How do other people react when it is suggested that they have gotten certain advantages in life because they are white?"*	
Part VI	The session leader discusses with a second actor in the room whether we can do things to counteract the injustices present in society. Participants are asked to propose solutions and concrete things that they could do to work against injustice. The suggestions made are written down and displayed on a screen in the room in real-time. Lastly, the participants are asked whether they will commit to their proposed solutions, and if they will sign the "new social contract" that they came up with together.	Evoking values and committed action to concrete examples of behaviours.
Ending	The session leader debriefs the participants on the purpose of the session and thanks them for their participation.	Debriefing and closing the session.