Preventing racial and religious hate

An evidence-based toolkit for preventing prejudice and vilification

	1300 292 153
Contact us	1300 891 858
Enquiry Line	
Fax	
NRS Voice Relay	1300 555 727 then quote 1300 292 153
Interpreters	1300 152 494
Email	enquiries@veohrc.vic.gov.au
Website	humanrights.vic.gov.au

Accessible formats
This document is available for downloading from our website at humanrights.vic.gov.au/resources in PDF and RTF. Please contact the Commission if you require other accessible formats.

Privacy
The Commission complies with Victorian privacy laws and the confidentiality provisions of the *Equal Opportunity Act 2010* (Vic). Our privacy policy is available online at humanrights.vic.gov.au/about-us/policies/privacy or by contacting us.

Disclaimer
This information is intended as a guide only. It is not a substitute for legal advice.

The Victorian Equal Opportunity and Human Rights Commission acknowledges that we work on the traditional lands of the Wurundjeri people of the Kulin Nation. We also work remotely and serve communities on the lands of other Traditional Custodians. We pay our respects to their Elders past and present.

Artwork: Madison Connors (2023), *Gorakor Wunbuni Yingurni ('Walk gently today')*

**Victorian Equal Opportunity
& Human Rights Commission**

Preventing racial and religious hate

An evidence-based toolkit for
preventing prejudice and vilification

July 2024

Acknowledgements

The Victorian Equal Opportunity and Human Rights Commission acknowledges Aboriginal people as the First Peoples, Traditional Owners and Custodians of the land and waterways, upon which our lives depend.

Our office is located on the lands of the Wurundjeri People – the lands on which many of our team members also live and work from home. We acknowledge and pay our respects to ancestors of this country and Elders past and present. We extend that respect to all First Peoples.

We recognise that First Peoples are disproportionately affected by prejudice and hate. In moving forward to address these issues, it is important to acknowledge the ongoing harms that colonialisation causes to First Peoples communities, including entrenched and systemic racism.

We honour the tireless efforts of generations of First Peoples who have stood against racism and advanced the values of freedom, fairness and equality. It is through the ongoing efforts of First Peoples leaders and communities that language, lore, custom and culture continue and flourish.

We commit to working with First Peoples to try to prevent historical injustices from continuing, including through truth-telling and developing Treaty.

Project acknowledgements

We would also like to recognise the knowledge and expertise held within Victoria's many communities on prejudice, vilification, prevention and drivers. In undertaking this project, the Commission has aimed to look beyond traditional academic research to understand the knowledge held and shared by communities.

We wish to acknowledge the contributions and insights of the researchers, community members and knowledge-holders – many of whom brought both research expertise and lived experiences in their perspectives – who generously participated in the project's consultations in challenging times. In particular, the Commission would like to thank the experts who provided critical reviews of this report:

- Associate Professor Matteo Vergani from Deakin University
- Dr Nicholas Faulkner from the Victorian Department of Premier and Cabinet's Behavioural Insights Unit and BehaviourWorks Australia at Monash University
- Dr Wing Hsieh.

Contents

1.
Introduction

Prejudice – negative attitudes or biases against a group of people – and **hate conduct** – behaviours and speech which express, support, cause or encourage prejudice – harm people, communities and societies.

In its preamble, Victoria's key law against vilifying conduct, the *Racial and Religious Tolerance Act 2001*, describes hate conduct as diminishing the dignity, self-worth and belonging of First Peoples, racialised people and marginalised religious people.[1] Decades of research show that prejudice and vilification harm health, lead to violence and marginalise communities.[2]

> They dissect our communities and degrade them to a point where communities feel really disempowered.
>
> — Consultation participant

As part of its response to the Victorian Legislative Assembly Legal and Social Issues Committee's *Inquiry into anti-vilification protections* (the Inquiry), the Victorian Government funded the Victorian Equal Opportunity and Human Rights Commission (the Commission) to undertake research into what drives – and what can prevent – racial and religious prejudice and hate in Victoria.

This report explores a range of drivers and influences that lead to prejudice and hate. It also provides an evidence-based toolkit of prevention approaches. These different approaches can be combined, adapted and extended – by governments, organisations, groups and individuals – in different ways using the best practice principles outlined in this report.

In preparing this report, the Commission analysed 167 items of research and evaluation from 2000 to 2024, spoke with leading experts in prejudice and hate, and heard from members of First Peoples communities, racialised communities and marginalised religious communities through a series of consultation sessions and a statewide survey. Community members spoke about how they continue to experience significant and harmful incidents of racial and religious hate, and raised concerns about incidents targeting children and families.

> It exists even right up to the highest offices in the country. It slips out – that's how we know it exists.
>
> — Consultation participant

Throughout the Commission's consultations, community members called for prevention of prejudice and vilification across many areas of life: education, workplaces, online, healthcare, public transport and even within families. Well-designed and evidence-based prevention, particularly strategies aimed at children and young people, have the potential to make positive direct and indirect impacts across Victoria.

However, despite a clear desire and need for prevention strategies, community members also expressed cynicism and frustration in relation to initiatives and structures aiming to reduce prejudice and prevent vilification.

> If nothing is happening, then I kind of lose faith in all these commissions and authorities and laws. If there's no implementation, are we just talking and talking and talking? We're not doing anything.
>
> **— Consultation participant**

To give prevention strategies the best chance of success and lasting impact, they need to be designed and implemented with **evidence**, **rigour** and **sustainability**.

Experts spoke about how past prevention strategies have been limited by short-term funding, short-term monitoring and evaluation (if any) and small-scale implementation. Scalability can be a challenge for some prevention approaches,[a] but it may be key to society-wide change. If a prevention strategy reaches enough people and groups, society may then reach a positive tipping point in shifting norms, attitudes and behaviours.[3]

As one expert noted, the scale of the issue links to the scale of the response needed. Prejudice is a large-scale issue that exists across systems and populations. In recent years, researchers have also begun applying a systems thinking approach to systemic prejudice.[4] This approach provides a fresh framework for understanding how prejudice arises from, influences and reinforces systems. It might also help reveal opportunities for communities and governments to disrupt systemic prejudice.

2024 marks 70 years since the publication of Gordon Allport's *The Nature of Prejudice*, a landmark work which shifted understandings of the origins of prejudice, its links to contact between groups, and how prejudicial behaviours can escalate from hate speech and discrimination to hate crimes and genocide.[5] In the decades since the book's publication, researchers and communities have continued to pursue better understandings of what works to eliminate prejudice. This report summarises what we know now about prejudice and vilification, and provides a toolkit to help the Victorian Government and community make the most of this knowledge.

---

a   Notably, Victorian researchers recently designed a scalability assessment framework for prejudice reduction interventions – see Wing Hsieh, Rebecca Wickes and Nicholas Faulkner, 'What matters for the scalability of prejudice reduction programs and interventions? A Delphi study' (2022) 10(1) *BMC Psychology* 107.

2.

About this report and the evidence

In 2020, the Victorian Legislative Assembly Legal and Social Issues Committee examined current anti-vilification laws and the effectiveness of the operation of the Racial and Religious Tolerance Act. The committee considered a range of issues, including evidence of increasing vilification and hate behaviours in Victoria.

The committee made 36 recommendations to strengthen Victoria's anti-vilification protections, including recommendation 3:

> That the Victorian Government fund ongoing research on the drivers behind vilification conduct and prejudice, and effective strategies to prevent this conduct.[6]

In 2023, the Victorian Government funded the Commission to prepare this research report in response to recommendation 3. The report focuses on racial and religious prejudice and vilification. It draws from a range of data sources to identify:

- **best practice principles** of how to design and implement prevention strategies
- the **drivers** behind prejudicial attitudes and vilification
- what empirical evidence[b] tells us about effective **prevention** approaches
- **case studies** of past prevention initiatives, including ones that combine prevention approaches.

There are commonalities between drivers and prevention of racial and religious prejudice and vilification and other forms of prejudice and vilification.[c] However, prejudice and vilification on the basis of other attributes (for example, disability, age, LGBTIQA+ status) can take different forms and have different characteristics.[7] There is the potential for further research in the style of this report focused on other forms of prejudice and vilification, as well as intersections between these forms.

2.1 About the Commission

The Commission is Victoria's human rights regulator. We are an independent body with responsibilities under 4 laws:

- the *Change or Suppression (Conversion) Practices Prohibition Act 2021* (Vic)
- the *Charter of Human Rights and Responsibilities Act 2006* (Vic) (the Charter)
- the *Equal Opportunity Act 2010* (Vic)
- the Racial and Religious Tolerance Act.

The Commission's role is to: protect and promote human rights; to eliminate discrimination, sexual harassment and victimisation to the greatest extent possible; and to prevent and respond to change or suppression practices. We do this through a range of functions, including: resolving complaints and reports; undertaking investigations; developing and delivering educational resources and programs; and assisting Victorians to understand their rights.

---

b   Empirical evidence is non-theoretical evidence that comes from observations, experiments or experiences.

c   This report draws from both evidence on prejudice and vilification in general and evidence specifically on racial and/or religious prejudice and vilification.

2.2 Definitions and terminology

This report uses terms drawn primarily from Victorian and international human rights frameworks, as well as literature on prejudice and vilification.

Charter of Human Rights and Responsibilities (the Charter)	The Charter is a Victorian law that sets out rights, freedoms and responsibilities for all people in Victoria. It is about the relationship between government and the people it serves. The Charter requires public authorities, such as Victorian state and local government departments and agencies, and people delivering services on behalf of government, to act consistently with the human rights in the Charter.
Drivers	These are the factors, influences and determinants behind an attitude or behaviour,[8] in other words, 'Why does someone think in a particular way?' and 'Why does someone behave or act in a particular way?'
First Peoples communities, racialised communities, and marginalised religious communities	These are the communities most likely to experience vilification on the basis of their race and/or religion. This is the result of negative racialisation and/or marginalisation on the basis of religion. We distinguish First Peoples because of their unique experiences of racism due to the colonial history of Australia. Language in this space is continually evolving and there are limitations to every term, and we use these terms based on current research, community preferences and the context of this report.
Groups	These are ways of classifying or categorising oneself and others, usually based around identity. A person's 'ingroup' is made up of people they perceive to be like them in a particular way (for example, having the same cultural background, working in the same job, supporting the same causes or political parties). An 'outgroup' is made up of people they do not identify with or perceive to be unlike them.
Perpetrator	This is a person who has vilified another person.
Prejudice	This is a negative attitude towards and bias against a particular group of people. Prejudice has multiple aspects: beliefs, emotions, values and intentions.
Racialisation	This is when people are categorised based on race. Racialisation is one way that the powerful people and organisations which control resources and influence systematically exclude minorities in order to make themselves more powerful. It is the basis on which hierarchies are structured and racism occurs, where some groups are racially privileged (or positively racialised) and others racially marginalised (negatively racialised). Throughout this report, when the term 'racialised' is used we mean 'negatively racialised'.

Scalability	This is the 'capacity of programs and interventions to increase in reach and impact'.[9]
Target	This is a person who has been subject to vilification on the basis of their (perceived or actual) race or religion.
Trigger events	These are events or incidents that trigger a broad community, societal, political or economic response.
Vilification (in this report)	This refers to **non-violent hate behaviours**: words, symbols, depictions or actions which express, support, cause or encourage prejudice or hate on the basis of someone's race or religion. Note that there is no standard definition of hate behaviour or vilification in law, research or advocacy in national or international contexts.[10] To consider the broadest possible range of evidence, this report **uses a broader definition of vilification than is in Victorian law**.[11]

2.3 Methods

The project used a multi-method qualitative design to consolidate existing empirical knowledge and evidence on the drivers and the effective prevention of racial and religious prejudice and vilification in Victoria.

Research scope	The project focused on 2 research questions: 1. What are the drivers (including influences, risk factors and protective factors) of racial and religious prejudice and vilification in Victoria? 2. What racial and religious prejudice and vilification prevention strategies have been researched or evaluated that could be implemented or adapted for the Victorian context? The project scope excluded some related matters (except where they overlapped with racial or religious prejudice or vilification – for example, hate crimes often include a hate speech element[12]): ▪ online aggression (including cyberbullying and trolling) without a vilification element ▪ discrimination (unfair treatment, practices or conditions) without a vilification element ▪ hate crimes (crimes against an individual, a group or property motivated by hatred, prejudice or hostility) ▪ prejudice and vilification based only on other attributes (noting that prejudice and vilification can be intersectional).
Research data sources	The project draws from a range of data sources to answer the 2 research questions, including: ▪ local and international **academic** and **grey literature**[d] from May 2000 to February 2024 – identified through targeted searches, snowballing,[e] suggestions from expert participants and the use of <u>Citation Tree</u> and <u>Connected Papers</u> ▪ **expert interviews** with 11 researchers, academics and other knowledge holders with expertise in behavioural insights, cultural studies, hate behaviours, online safety, prejudice interventions and radicalisation – identified through the Commission's networks, a review of authors of identified literature and suggestions from expert participants ▪ **group discussions** with a range of community members: — 1 with First Peoples who responded to a callout to participate in an in-person session facilitated by the Victorian Aboriginal Legal Service — 4 with members of the Victorian Multicultural Commission (VMC) Regional Advisory Councils who responded to a callout to participate in a series of online group discussions co-hosted with the VMC — 1 with dispute-resolution staff at the Commission who have worked on vilification complaints under the Racial and Religious Tolerance Act ▪ 462 responses from Victorian community members to a public **online survey** on racism and religious hatred, available in 9 languages.

---

d   Grey literature is literature outside academic journals and publishers. It includes reports and publications from governments and non-government organisations.

e   Snowballing refers to reviewing relevant references in identified literature to identify more literature.

	Collection of qualitative data focused on the 2 research questions, as well as additional questions on contexts and settings of vilification in Victoria. To complement this, the project also considered relevant **quantitative data** on: ▪ enquiries, reports and complaints to the Commission under the Racial and Religious Tolerance Act between 1 July 2013 and 31 December 2023 (covering the last 10 financial years of data, plus the full 2023 calendar year) ▪ reports to Victorian and Australian registers of racial and religious discrimination and vilification.
Included literature	We identified 315 relevant items of literature. Of these, we analysed 167 items that met our core inclusion criterion of empirical (non-theoretical) evidence that helped us answer the research questions.[f] The analysed literature consisted of: ▪ 24 meta-analyses or systematic reviews ▪ 1 scoping review ▪ 5 randomised control trials ▪ 33 quasi-experimental studies ▪ 85 case, cohort, cross-sectional, longitudinal, post-test only, pre- and post-test, or qualitative studies ▪ 15 narrative reviews or literature reviews ▪ 4 expert opinion or Delphi studies. Of the 167 analysed items, 47 looked at prejudice or hate in Australia (21 of these focused on Victoria). Research in this space has increased significantly in recent years, with 106 of the 166 items published between 2020 and 2024. All analysed journal articles also underwent brief credibility checks using **Beall's List**, **PubPeer** and **Retraction Watch**, as well as searches of the **Directory of Open Access Journals** and use of a **journal evaluation tool** in some instances. The publishers or journals for 7 items were flagged as having potential quality or credibility issues – we still included these items; however, we treated the strength of their evidence as on par with that of non-peer reviewed literature items. While this report captures as much relevant evidence as possible, it does not comprehensively cover the full breadth of empirical prejudice and vilification literature. Research, evaluation and practice are ongoing in this space.

---

f   While all 166 items were analysed, not all are directly cited in this report. See **Appendix** for the full list.

Excluded literature	We excluded 148 items of literature for one or more of the following reasons – an item:
	▪ did not include empirical research into the drivers of prejudice or vilification
	▪ did not evaluate or test the effectiveness and/or impact of one or more prevention initiatives
	▪ did not contribute to our understanding of drivers or prevention approaches
	▪ had insufficient relevance to the Victorian context (for example, items that only covered hate speech in the context of violent civil conflicts overseas)
	▪ covered *only* related but out-of-scope conduct (for example, online aggression, discrimination, extremism, hate crime, radicalisation, terrorism).
How the data was analysed	This report is primarily a qualitative synthesis of the included data sources. Our analysis focused on identifying:
	▪ empirical evidence and community understandings of drivers
	▪ empirical evidence and community perceptions of prevention strategies
	▪ conditions that may increase prevention strategy effectiveness
	▪ potential challenges in implementing prevention strategies
	▪ contexts of concern for vilification in Victoria.
Data that was not analysed	We did not conduct any quantitative analysis. This report does not include any statistical analyses or reporting of effect sizes of prevention strategies (for example, percentage improvements in particular attitudes or intentions).

2.4 The Victorian context

Community members told the Commission that hate incidents have increased in frequency and severity in waves linked to major events (see **4.4 Media narratives and trigger events**) – most recently, during the COVID-19 pandemic and associated restrictions, before and after the Voice Referendum, and through the conflict in the Middle East starting in 2023. This is supported by reports of and enquiries about racial and religious vilification to the Commission:

▪ More than one-fifth (22.6%, n=353) of all racial and religious vilification enquiries made in the 10-year period up to 30 June 2023 occurred in the first 12 months of COVID-19.[g]

▪ From 1 January to 31 December 2023, 42% (n=37) of racial and religious vilification enquiries occurred between October and December, coinciding with the outcome of the Voice Referendum (14 October) and the beginning of conflict in the Middle East starting in 2023 (7 October).

---

g   March 2020 to February 2021.

Understanding reported and recorded data

There are limitations to reported and recorded data about vilification and other hate behaviour. Under-reporting is an ongoing issue locally and internationally for a range of reasons, including low awareness, fear of consequences, low trust in authorities and low accessibility.[13] If a report is made, the recipient first decides whether and how to record it, and then how to analyse or release reporting data.

Report statistics might not always accurately reflect incident trends because they can be influenced by several external factors, for example:

- communication and engagement campaigns (for example, campaigns to raise awareness of reporting options)

- encouragement or discouragement within communities

- changes in trust towards recipient organisations about how they might respond to or handle a report

- perceived lack of outcomes following past reports (either an individual's own reports or other people's reports)

- changes in recording practices within recipient organisations (for example, if recipient organisations change their thresholds or definitions of hate conduct)

- changes in individuals' and communities' perceived thresholds for behaviour (for example, if more people recognise certain conduct is vilifying)

- how recipient organisations choose to analyse, aggregate and/or release data.

Readers of these statistics should remember that there is likely a larger picture of vilification which has not been reported to or recorded by authorities or registers.[14]

Community members spoke about many contexts of concern where they have experienced and/or witnessed vilification:

- schools and universities
- workplaces and volunteering
- social media and online comments
- news media and political discourse
- sports and gaming venues
- civic participation (for example, at council meetings and fora, on local government social media)
- healthcare and support services (for example, interactions with hospital staff, general practitioners, paramedics, aged care services)
- public transport
- extended families (for example, comments and behaviour from in-laws, relationships being forbidden on the basis of race or religion)
- general public life (for example, interactions at petrol stations, retail outlets, parks, restaurants, bars, local events, social events).

> Wherever racialised Victorians can be found, there will be racism. It just depends on what form that takes.
>
> **— Consultation participant**

Several of these areas are also reflected in the number of formal complaints of racial and religious vilification made to the Commission's dispute resolution service. Between 1 July 2013 and 30 June 2023 (a 10-year period):

- 50.3% (n=81) of complaints were about vilification in general public life – in public places, at social events, or in hospitality or retail
- 17.4% (n=28) were about vilification in media, press or advertising
- 13.7% (n=22) were about vilification in social media – this is notable given the additional barriers in raising complaints against international social media platforms (for example, identifying and engaging with staff in overseas organisations like Meta/Facebook).

2.5 Caution in interpreting findings

A core finding of our research is that the existing body of knowledge and evidence on preventing prejudice and vilification is inconclusive and can only indicate what might work. While some prevention approaches have promising results, they have often not been tested, replicated or evaluated at larger scales or across more diverse populations.[15] Interventions to reduce prejudice and vilification are no exception.[16]

Studies in this field must be read and interpreted with caution, even if a study's findings are promising (see also **3 Best practice principles**). When applying evidence to other cohorts and contexts, readers should think about potential considerations and limitations of the research, including:

- study design or appropriateness (for example, presence of adequate control or comparison groups, sample sizes)
- participant selection and appropriateness (for example, how participants were recruited, demographics or homogeneity of participants, whether the experiences and effects for the participants can reasonably be generalised to larger and more diverse populations)
- outcome measure design or appropriateness (for example, selection of outcome measures, what they measure, whether they are validated measures, whether they rely only on self-reporting of attitudes or behaviours, whether they focus on behavioural intentions and not actual behaviours)
- timing of and effects found in measures or tests (for example, if and when pre-tests and post-tests were administered, how many pre- and post-intervention tests were administered, how long the follow-up was, whether effects changed or were stable over longer term follow-ups)
- setting of a tested intervention and how close it was to a real-world context (for example, a laboratory or a real-life setting)
- context of an intervention (for example, what country it was tested in, how long ago the research was undertaken, changes in the sociopolitical and economic environment)
- replication of effects across multiple trials of the same or a similar intervention
- scalability and adaptability of an intervention.

These factors may mean that an intervention might not have the same results when replicated. For example, a 2022 meta-analysis identified a potentially promising prejudice-reduction approach called perceived variability (see **5.4 Perceived variability**) based on encouraging results from predominantly French studies.[17] A follow-up trial in Australia was unable to replicate these effects.[18] As one 2021 review of prejudice-reduction experiments commented:

> One urgent message of this article is the need for caution when reviewing prejudice reduction recommendations generated by small studies with optimistic conclusions.[19]

In this report, the Commission has adopted a cautious approach to the evidence available. More research and evaluation are needed to build on the existing evidence base.
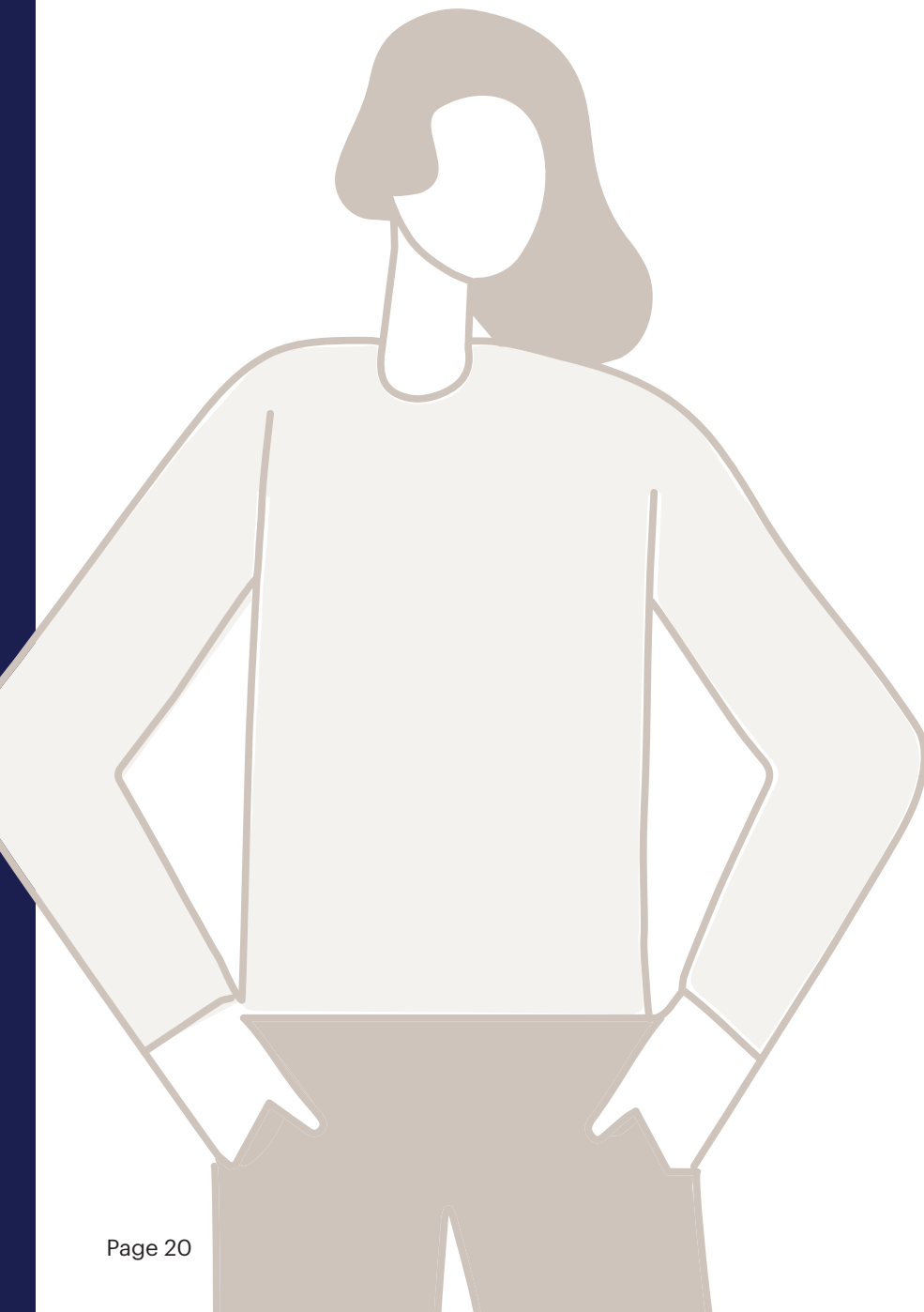
3. Best practice principles

Strategies to prevent prejudice and vilification can take different forms. As shown later in this report (see **6 Case studies**), they can focus on a single prevention approach or combine multiple approaches (see **5 Prevention**). They can also aim to address drivers at broader social or systemic levels or at a more individual level (see **4 Drivers**).

For prevention strategies to have a greater chance of success and lasting impact, they should be designed based on evidence, be implemented with rigour and aim to be sustainable. From our consultations and research, the Commission has identified 9 key best practice principles that prevention strategies should aim to integrate:

1. Identify the **evidence-based change mechanisms[h] and outcomes** the initiative will target[20] – and combine multiple prevention approaches where possible to maximise chances of effectiveness.[21]

2. Design and tailor strategies to the **social, political, legal and cultural contexts**.[22]

3. Consider **scalability** when selecting which approach(es) to use and to ensure the best possible reach of an overall prevention strategy.[i]

4. Engage with and understand the relevant research using a cautious and critical eye in order to **align an intervention with the evidence**, **address past limitations** and **integrate lessons learned** (see also **2.5 Caution in interpreting findings**).

5. Use **community-centred and -led design approaches** (for example, Design Justice, Equity-Centred Community Design) where possible to meaningfully engage with communities, their contexts and their knowledge. These approaches can help tailor interventions to new contexts, encourage acceptability of interventions in communities and ensure that interventions minimise risks of cultural load and unsafe environments.

6. Consider engaging and integrating **behavioural change design, implementation and testing expertise** (for example, through the Victorian Government's Behavioural Insights Unit).

7. Set up outcome-focused **monitoring and evaluation frameworks and mechanisms** early[23] – and strengthen these with rigorous methods and validated measures where possible.

8. Use **pilot tests and trials** to check for intended, unintended and undesired outcomes[24] (for example, potential backlash, defensiveness, perceived condescension or other counterproductive effects).

9. **Share knowledge and lessons learned** from pilot tests, research, process evaluation and outcome evaluation with communities and researchers to help future prevention strategies maximise their effectiveness, know what has been done and avoid what has not worked.

---

h   Change mechanisms are the underlying ways in which an intervention is expected to change an attitude or behaviour.

i   For example, using the scalability assessment framework developed by Wing Hsieh, Rebecca Wickes and Nicholas Faulkner, 'What matters for the scalability of prejudice reduction programs and interventions? A Delphi study' (2022) 10(1) *BMC Psychology* 107.

4.
Drivers

Attitudes (like prejudice) and behaviours (like vilification) are the complex result of many types of interrelated and interacting drivers, including influences, determinants, protective factors and risk factors. There are internal psychological drivers, social and relational drivers, and structural and environmental drivers.[25]

In the local and international evidence, some drivers are more frequently identified and explored than others:

- feelings of threat
- media narratives
- trigger events
- exposure to hate (particularly online).

Some of these were also raised by expert interviewees and community members based on their experiences and observations.

What are trigger events?
Trigger events are events or incidents that trigger a broad community, societal, political or economic response (for example, terrorist attacks, violent crimes, economic crises, pandemics). Notably, these triggers can involve either factual events or instances of misinformation.   Other triggers for hate can include political debates and events or celebrations associated with particular cultures, ethnicities or religions.

In both the consultations and the analysed literature, experts noted that even where all possible drivers and risk factors are present, this does not guarantee that someone will be prejudiced or will vilify – although it would increase the likelihood.[26]

4.1 Understanding the range of drivers

Everyone consciously and unconsciously holds attitudes, ways of thinking and feeling, cognitive biases, and mental shortcuts. These can include negative beliefs about or intentions to inflict harm on people from another group.
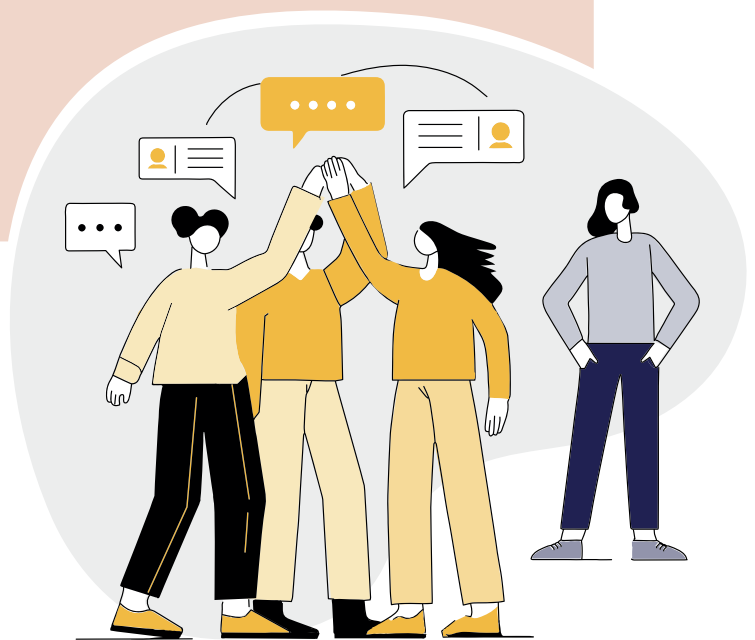
What are groups?

Ingroups and **outgroups** are ways of classifying or categorising oneself and others, usually based around identity. A person's ingroup is made up of people they perceive to be like them in a particular way (for example, having the same cultural background, working in the same job, supporting the same causes or political parties). An outgroup is made up of people they do not identify with or perceive to be unlike them.

Intergroup attitudes, feelings and interactions are those between different groups. Some intergroup behaviours can be harmful or hostile (for example, vilification).

Some interventions focus specifically on interactions between **majority groups** and **minority groups** – these terms refer to groups' status within a society and its demographics (for example, in Australia white Australians are a majority group and Chinese Australians are a minority group).

Attitudes are hard to measure. They can change over time, and it is often difficult to trace their origins.[27] Attitudes, mindsets and biases are all important in driving vilification. Recent research has looked into how strongly attitudes and social norms predict a person's intent to vilify. These studies found signs that attitudes are a better predictor than social norms.[28]

What are social influences and norms?

Social influences refer to how other people and groups influence a person's attitudes and behaviours. These include **norms**, which are perceptions about what people feel expected to do, what they believe others in their group think and how they believe others act.[29]

These norms – combined with social pressures, power dynamics and group cohesion – can shape how people think about and behave towards other people, both inside and outside their groups.

4.2 When prejudice starts

Prejudice and its drivers exist across individual, social and structural levels.[30] Evidence shows that prejudice development starts in early childhood,[31] before appearing to level out as children reach their early teen years. Over a life span, different drivers shape people's attitudes to different extents, including:

- transmission of racial and/or religious prejudice from parents and within families[32]

- desires to fit in or share 'desirable' views with others (for example, friends, classmates, co-workers, social groups)[33]

- emotional responses to news or events[34]

- what a person values or cares about, and others who share those values, like their perceived ingroup[35]

- ideological predispositions towards societal hierarchies, social dominance and mindsets that see the world as dangerous, threatening and competitive – these are generally measured using the Right-Wing Authoritarianism and Social Dominance Orientation scales[36]

- indirect exposure via communications and media, stories and narratives, and second-hand experience.[37]

Community members and experts also pointed to the historical roots and contemporary forms of racism, religious discrimination and hate as driving and reinforcing prejudice.

The Commission heard about how historical prejudice and exclusion continue to be reflected in today's society. One community member described the continued impacts of colonisation and marginalisation in Australia as being 'in the fabric of the country'. These range from the impacts of government policies and practices on First Peoples (for example, dispossession and removal from Country, removal of children)[38] to the consequences of terrorism-related narratives for Muslim communities[39] (see **4.3 Feelings of threat**).

4.3 Feelings of threat

Fear of others is an important driver of prejudice. The intergroup threat (or integrated threat theory) model describes 4 elements of threat that contribute to the development of prejudice:[40]

Intergroup anxiety	When someone feels anxious during an intergroup interaction about possible embarrassment, rejection or ridicule from others
Perceived realistic threats	Threats thought to be posed by an outgroup to an ingroup's physical wellbeing and safety, or access to or competition for resources (like housing or jobs) or political and economic power
Perceived symbolic threats	Threats thought to be posed by an outgroup to an ingroup's values, identities, norms or beliefs (for example, national or cultural identities and values, religious beliefs)
Negative stereotypes	Stereotypes and generalisations about an outgroup, including expectations of their characteristics or behaviours

These elements are all linked and can interact with each other. They can also exist at individual and group levels, in attitudes, behaviours, social norms and narratives.[41] The development of intergroup threat can be influenced by a range of factors, including:

- the way someone thinks and processes information (for example, rigid binary thinking, filtering information in a way that supports their existing opinions)

- their own direct or observed experiences with an outgroup

- narratives from the media or other people that promote feelings of threat, fear, social incompatibility or differences in values (see **4.4 Media narratives and trigger events**).[42]

Each element of intergroup threat appears to affect prejudice and vilification differently (see **4.5 When prejudice becomes hate**). Research indicates that perceived realistic threat and symbolic threat are generally better predictors of prejudice and vilification than is intergroup anxiety.[43] For example, economic insecurity or instability can lead to perceived intergroup competition over access to jobs or housing, with some ingroup members blaming outgroups for the instability. Research from the USA and Italy supports the existence of links between economic insecurity or instability and rises in racial prejudice[44] and online hate.[45]

Similar links between perceived threat and terrorism concern have been established, particularly in communication and media contexts. People's perceptions of threat and safety, particularly after trigger events, are influenced by how media stories are framed, the content of narratives and the presentation of online discussions (see also **4.4 Media narratives and trigger events**).[46] Evidence indicates that those with higher concern around terrorism are more likely to hold anti-Muslim views (regardless of their level of factual knowledge of Islam) and that these views affect what information they can recall about Muslims.[47]

4.4 Media narratives and trigger events

The media landscape, particularly news media, plays a key role in developing, spreading and reinforcing prejudice.[48] Both online and traditional media outlets provide, frame and spread information. This provides opportunities to influence or persuade people about what they think, how they act and how they respond to information.

Media reporting can directly and indirectly present and reaffirm existing negative stereotypes and ideas about specific outgroups across the community.[49] This happens both intentionally and unintentionally when media outlets:

- provide platforms to individuals or groups with prejudicial views[50]
- create cultural or social narratives out of unique individual incidents[51]
- stir up moral panic by treating individual actions as representative of a community[52]
- spread or reinforce misinformation or disinformation[53].

The way media outlets respond to trigger events is also impactful on creating or reaffirming prejudicial views.[54] These views are often negative responses from the majority group towards a minority group.[55] Geopolitical conflict beginning between groups, terrorist attacks, pandemics and local crime incidents are some of the commonly identified types of trigger event that can influence prejudice and vilification.[56] Notably, there are indications that events or celebrations associated with particular cultural, ethnic or religious communities can be neutral triggers for non-violent hate behaviours.[57]

The initial effects of a trigger event may be small scale and localised, or they may have large-scale impacts nationally or internationally. For a trigger event to have an effect on attitudes or behaviours, people need to know that it has happened. This often occurs first through reporting of an event or incident by news media, but it can also occur through political discourse or social media outside of traditional news outlets. Both research and experts recognise trigger events and the narratives surrounding them as notable drivers of prejudice and vilification.

The key factors that influence the prejudice-spreading potential of trigger events include:

- the framing of an event – and whether the messaging is negative[58]

- the presentation of facts or speculation – and whether they have been verified

- the portrayals of any actors or communities involved or potentially involved – and whether these portrayals racialise those involved[59] or differentiate broader communities from the actors directly involved[60]

- the potential for these factors to interact with existing or emerging community dynamics or norms.

These narratives and factors have real-world consequences for what people think and how they behave towards outgroups presented in the media. Media communication of trigger events can activate or increase perceptions of intergroup threat (see **4.3 Feelings of threat**), as well as people's perceptions of intergroup differences and the capability of different groups to safely live together.[61] This can ultimately lead people to feel justified in developing or reinforcing negative attitudes and mindsets about another group and, in some instances, vilifying members of that group.[62]

4.5 When prejudice becomes hate

There is no known 'formula' of drivers that produces (or prevents) a person becoming prejudiced or engaging in vilification. There are only factors that make prejudice and hate more (or less) likely to occur. However, the evidence demonstrates that hate behaviours have their origins in prejudice. For example, initial research to validate the Hate Behaviours Scale (HBS), a new measure for intention to engage in (non-violent and violent) hate, found a significant relationship between anti-Muslim prejudice and intention to engage in anti-Muslim hate behaviours.[63] Threat is an important element of this – some of the intentions measured by the HBS are about 'defending' an individual's community against the perceived threat posed by an outgroup (in this instance, Muslim people).

Both prejudice and intergroup threat (see **4.3 Feelings of threat**) are significant drivers of hate. Research has investigated how prejudice and threat can influence people's behaviour through:

- the ways an individual sees the world and their relationships to other groups, including neighbourhood safety and social cohesion[64]

- beliefs that outgroups are threats to social cohesion or to access to resources[65] (see also **4.3 Feelings of threat**)

- rejection of groups that an individual believes do not share the same values as them or threaten those values.[66]

Research suggests that perceived threat is more likely to result in confrontation than is intergroup anxiety.[67] This is consistent with recent research outside this field which has found that people anxious about terrorism are more likely to avoid violent terror content, whereas people angry about terrorism are more likely to seek and consume violent terror content.[68] Further research is needed into whether this effect extends to consumption of hate content, which is important because of racial and religious stereotypes regarding terrorism (see **4.4 Media narratives and trigger events**) and of the role that exposure to hate has in driving further hate (see **4.6 Exposure to hate**).

Vilification can also be driven by social influences, including normalisation of and desensitisation to stigma, marginalisation and discrimination.[69] Social influences can be exerted at structural and institutional levels, for example through social and political responses to large-scale trigger events like the September 11 attacks,[70] the COVID-19 pandemic[71] and the 2005 Cronulla riots.[72] These events drastically changed the nature of political and social norms about how Muslim, Arab and Asian communities are perceived, treated and expected to behave. They also changed how other groups see and interact with these communities, leading to increased hate.[73] Research is yet to be published on the full influence of the social and political response to the 2023 Australian constitutional referendum (on the Aboriginal and Torres Strait Islander Voice to Parliament), but researchers and community members have commented on how this response has similarly led to increasing vilification against First Peoples.[74]

Drivers of extreme attitudes, violent behaviours and hate crimes

The drivers described in this report overlap with commonly identified drivers of extremism and hate crimes.[75] While discussing these extreme attitudes and violent behaviours is out of the scope of this report, it is notable that:

- research indicates that exposure to violence, maltreatment and substance use are associated with extremist attitudes among adolescents[76]

- experts have noted associations between childhood stressors and trauma with the use of violent hate[77]

- early evidence suggests that misogyny is a predictor of violent extremist intention.[78]

4.6 Exposure to hate

People are increasingly exposed to hate speech[79] and online aggression.[80] Emerging evidence shows that repeated exposure to hate speech leads to:

- desensitisation to offensive and hateful content

- limited ability to empathise with either outgroup or ingroup members

- dehumanisation of groups targeted by the content

- greater distancing from the targeted group

- increased prejudice.[81] [j]

Evidence also suggests that people are more likely to make trolling or aggressive comments online if they are in a negative mood and if the discussion context already has a negative or aggressive tone.[82]

Increased exposure to hate speech and aggression online, combined with online anonymity, has implications for younger generations, who are increasingly introduced to online environments earlier in life. In general, adolescents are more likely to be exposed to hate speech online than to perpetrate it themselves.[83] In addition to this individual-level behaviour, people online are building communities of like-minded individuals who find legitimacy for and recognition of their prejudicial views, which can frequently spill over into the offline world (see **4.7 Online hate communities**).[84]

In recent years, evidence has emerged to support the existence of a mutual (and mutually reinforcing) relationship between online hate (for example, through social media, news article comments) and offline hate behaviours.[85] Recent research in the USA demonstrates that negative online speech from key authority figures about Muslim[86] and Asian communities[87] can alter social norms – including legitimising and destigmatising large-scale prejudice towards entire outgroups and emboldening others to express prejudicial views.

---

j   This is an important consideration for individuals, groups or organisations that repeatedly engage with hate content (for example, when monitoring, responding to or researching content). There is a need for available, accessible and appropriate support measures.

4.7 Online hate communities

Online spaces and communities play increasing roles in spreading and normalising vilification.[88] Social influences and norms significantly facilitate online hate and hate communities.[89]

Online environments are used differently by groups and individuals to achieve different objectives. Individuals look for community, shared beliefs and approval of their views. By contrast, groups want to grow, spread their beliefs, gain recognition and create a unified identity. There is evidence that online environments are increasingly used for recruitment and mobilisation, community building and shaping, and role modelling to others.[90]

While much attention has been given to online hate communities on social media platforms since the 2010s, this phenomenon is not new. For example, the 2005 Cronulla riots acted as a trigger event that contributed to increased engagement and stronger bonds within a major Australian online far-right community.[91]

In online settings, networks and communities with like-minded attitudes and behaviours can easily gather, with members reinforcing each other.[92] This can create a feedback loop, where people receive information that confirms their beliefs, which then influences what information they look for. It can also set strong social incentives for and against the use of online hate.[93]

5. Prevention

For several decades, researchers have investigated how to prevent prejudice and vilification effectively. Various approaches for interventions have emerged over this time, with new approaches continuing to be proposed, developed and tested.

Experts have reiterated in literature and during our consultations that interventions are more effective when their designers identify and understand the theory and evidence behind what they are trying to change and how the intervention will achieve this.[94]

There are various ways of categorising interventions approaches to prevent prejudice and vilification (for example, by setting, by target population, by format). This report explores 5 key approaches (tested in either real-world or laboratory settings) that appear frequently in the evidence:[k]

Awareness	Reducing prejudice and/or vilification through learning and reflection
Categorisation	Reducing prejudice by targeting how people categorise themselves and others
Contact	Reducing prejudice through different forms of (positive) interaction between members of different groups
Perceived variability	Reducing generalisations and stereotypes by demonstrating the diversity within other groups
Perspective-taking	Reducing prejudice by exchanging narratives and actively engaging empathy

These approaches are not mutually exclusive and prevention strategies often combine multiple approaches (for example, contact plus categorisation, extended contact followed by direct contact, perspective-taking plus awareness) in different contexts (for example, schools, workplaces, media, online). They are also not the only prevention approaches that have been developed or will be developed.

In addition to the above 5 key prevention approaches, this section summarises a range of other notable prevention approaches.

---

k   These categories are drawn from an existing prejudice-focused classification scheme based on the underlying theoretical foundations of different prevention approaches. See Wing Hsieh, Nicholas Faulkner and Rebecca Wickes, 'What reduces prejudice in the real world? A meta-analysis of prejudice reduction field experiments' (2022) 61(3) *British Journal of Social Psychology* 689.

5.1 Awareness

How the intervention works

Awareness interventions aim to reduce prejudice and/or vilification through learning and reflection – by improving an individual's knowledge and understanding of prejudice, vilification and harm, encouraging critical thinking, and building their skills and confidence. These are usually education programs (or program components) in schools or workplaces, or media campaigns.[95]

Awareness interventions may cover a range of topics, for example:

- concepts, understandings and origins of prejudice and hate (see **Case study 6.3**)

- cross-cultural knowledge, understanding or awareness

- what words or actions are prejudicial or harmful and why

- cognitive and social processes and drivers behind prejudice or vilification

- how to respond to or challenge prejudice or vilification (see also **5.6.4 Social norms** and **5.6.5 Prevention through response**)

- the roles that individuals, groups, organisations and institutions can play in reducing or preventing prejudice, discrimination or vilification.

As awareness is based around learning and reflection, the Commission notes and acknowledges the importance of racial literacy – as well as approaches which centre First Peoples, racialised and marginalised religious perspectives and experiences – as part of awareness interventions.[l]

Awareness interventions are relatively flexible and can be designed or delivered in multiple ways, including:

- on their own or in combination with other approaches (for example, contact, social norms)

- using different types or combinations of facilitators or messengers (for example, external facilitators, peer leaders, train-the-trainer models)

- with varying levels of engagement (for example, passive learning, active reflection, creative methods)

- as one-off, periodic or ongoing learning or reflection.

Depending on their design, awareness interventions have the benefit of scalability: they can reach larger numbers of people more easily than other more time- or resource-intensive interventions.
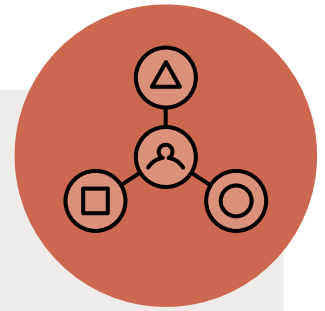
Notably, the Inquiry recognised the importance and potential impact of awareness interventions in **recommendation 7**, for community education campaigns on vilification and hate conduct.[96]

---

l   See the Commission's **Guideline: Race discrimination in the workplace** for more information about racial literacy.

How effective the intervention is	While awareness interventions are on average effective in real-world settings, their effectiveness can vary considerably depending on their design, delivery and content.[97] For example, direct attempts at persuading people to recognise and change their attitudes have the potential to backfire, and there have been past media-based awareness interventions which have inadvertently reinforced stereotypes.[98]
	A 2016 meta-analysis of 40 years of diversity training programs looked at a range of outcomes, including diversity-related *attitudes* and diversity-related *knowledge*. While these programs are generally effective on attitudes immediately after delivery, this effectiveness on attitudes decays over time. However, participants do retain knowledge over time.[99]
	A 2013 systematic review of school-based intercultural understanding programs found that building cultural knowledge and cultural awareness:
	▪ may result in positive short-term changes but not necessarily long-term changes in attitudes and behaviours
	▪ can have little if any effect on attitudes or behaviours
	▪ may backfire and reinforce prejudices if students' attitudes towards people from different cultural backgrounds are not explicitly addressed and thoughtfully discussed.[100]
	There has been no equivalent review for anti-racism education. Contemporary anti-racism education continues to evolve and there is a need for more evidence and rigorous evaluation around its design, delivery and effectiveness.
What factors increase its effectiveness	The key factors increasing the effectiveness of awareness interventions appear to be format, context and participants' pre-existing views. There is evidence that integrated approaches to diversity training (where the training is complemented by other initiatives) are more effective than standalone training, and that longer training is more effective.[101] A 2020 Victorian review of contemporary anti-racism strategies also noted the importance of anti-racism training:
	▪ being neutral, informal and respectful
	▪ providing factual information
	▪ combining multiple instruction and learning approaches
	▪ being tailored to the context
	▪ encouraging intergroup contact where appropriate
	▪ being delivered by trainers with lived experience as well as experience or qualifications in organisational change.[102]
	This 2020 review also noted the importance of communication- and media-based awareness interventions:
	▪ using explicit messaging
	▪ addressing specific negative beliefs instead of trying to create positive feelings
	▪ focusing on specific racialised groups rather than broad diversity or multiculturalism messages
	▪ highlighting commonalities between groups
	▪ pre-emptively addressing potential counter-arguments
	▪ considering the use of nonverbal behaviours (for example, posture, eye contact, personal distance) in visual media.[103]

	Recent research indicates that anti-prejudice messaging is more likely to be effective among individuals with lower levels of Right-Wing Authoritarianism (see **4.2 When prejudice starts**).[104]
	Although there is limited evidence on the effectiveness of communication- and media-based interventions in general, a 2007 review for the UK Department for Communities and Local Government drew from evidence available at the time to create 2 tools to help strengthen the persuasiveness and effectiveness of messages aimed at reducing racial prejudice:
	▪ a 27-item checklist for assessing initiatives[105]
	▪ key psychological mechanisms of messages – these include detailed guidance on a message's structure, its content and its delivery.[106]
What limitations and risks may be present	**Unintended negative consequences:** As noted, some awareness interventions have had unintended negative consequences (for example, unintentionally reinforcing stereotypes or biases, being perceived as condescending).[107]
	Limitations of knowledge on changing attitudes: Even if an awareness intervention successfully increases a person's knowledge of another group, this may not translate to reduced prejudice.[108] For example, research supports a link between terrorism concern and Islamophobia regardless of level of factual knowledge about Islam, indicating the limitations of interventions that only aim to improve knowledge.[109] More broadly, research into misinformation indicates that factual knowledge often fails to correct biased views.[110] Further research is needed to determine the impact of knowledge on changing behaviours.
	These limitations do not detract from the need for improved racial and religious literacy. In the Commission's consultations, many community members supported the role of education in reducing prejudice and increasing awareness.

5.2 Categorisation

How the intervention works	Categorisation interventions are based on cognitive and social categorisation theories about how people classify themselves and others into ingroups and outgroups.[111] There are different types of categorisation interventions, each of which aims to reduce prejudice in different ways: ■ **decategorisation**, where group memberships are de-emphasised and individual identities are emphasised ■ **common ingroup identity categorisation**, where the ingroup and outgroups are recategorised as belonging to a new overarching (superordinate) identity category – this superordinate identity becomes a new common ingroup, intended to extend ingroup favouritism to all members ■ **dual identity categorisation**, where both the original group identity and the new superordinate identity are simultaneously engaged.[112] Decategorisation and common ingroup categorisation share similarities with 'colourblind', 'colourmute' and 'racelessness' approaches to prejudice reduction, which attempt to avoid seeing, acknowledging or discussing race or racial difference.[m] The assumption is this will prevent racial bias; however, evidence indicates these approaches are counterproductive and can actually reinforce prejudice and stereotypes.[113]
How effective the intervention is	Categorisation-only interventions have not been frequently tested in real-world settings but, based on limited studies, they seem to be less effective on average than contact and awareness approaches.[114] Common ingroup identity and decategorisation interventions are ineffective and may have negative effects among children and adolescents, although researchers have not empirically determined why.[115] However, there is evidence supporting the effectiveness of dual identity categorisation (see below).
What factors increase its effectiveness	There is emerging evidence that dual identity categorisation interventions can be effective with children and adolescents when combined with virtual contact[116] or extended contact[117] approaches (see **Case study 6.1**).
What limitations and risks may be present	**Resistance:** One expert participant noted that with common ingroup categorisation interventions, they had observed a risk of resistance or backlash to a new superordinate identity being imposed. They also noted that when exposed to a recategorisation intervention, which explicitly draws participants' attention to ingroup and outgroup memberships, some participants instead focus on their own group as being more important.

---

m   We recognise these terms have connotations referring to disability and disabled communities. We use them here in the same way as the literature to discuss prejudice-related phenomena.

Risks to children: Categorisation interventions aimed at children may risk introducing social categories and ideas of difference earlier than they would have otherwise been introduced, potentially sowing the seeds for prejudice earlier in life.[118] Public authorities implementing categorisation interventions involving children should consider whether the intervention engages the right under the Charter to the protection of children as is needed in their best interests and whether there is any risk that this right might be unjustifiably limited by the intervention.[119]

Erasing identities: As with 'colourblind' and similar approaches, decategorisation and common ingroup categorisation approaches have a risk of undermining, dismissing or erasing cultural identities. These approaches also do not acknowledge that minority members generally do not get the choice to de-emphasise or ignore group memberships, as they are subject to society's ongoing racialisation of themselves and their communities.

Notably, dual identity categorisation interventions do not face these same limitations.

5.3 Contact

How the intervention works

Contact interventions are based on Allport's contact theory, which has been a key theory in understanding and reducing prejudice since the 1950s.[120] The core idea is that positive interactions (contact) between members of different groups will improve perceptions and attitudes between the groups, reduce favouritism or bias towards one's own group, and reduce intergroup prejudice or tensions.[121]

There are 2 main types of contact – **direct** and **indirect** contact – which can generally be split into 5 approaches:

- **contact meetings (direct)**, where participants from different groups engage in structured intergroup discussions that explicitly address intergroup relations

- **cooperative learning programs (direct)**, where participants from different groups work towards a common learning objective or product not related to interethnic relations (and intergroup relations are not discussed)

- **extended contact (indirect)**, where there are friendships or positive relations between a member of a person's ingroup and an outgroup member (or positive intergroup relations are portrayed in the media)

- **virtual contact (indirect)**, where participants from different groups engage in computer-based or online systematic discussions or activities

- **imagined contact (indirect),** where participants experience a simulated positive interaction with an outgroup member (notably, while imagined contact can take place through media like television and film, imagined contact research tends to be laboratory-based).[122]

There are also other forms of extended and indirect contact in the literature, including representations or portrayals in the media (sometimes called parasocial contact or media contact), cultural events and conversations or interactions.[123] Some of these can include an element of perceived variability (see **5.4 Perceived variability**).[124] Although these other forms of contact do not meet the original optimal conditions of contact (see below), there is increasing research into these approaches and their effectiveness.

Contact interventions engage multiple processes, including:

- experiencing exposure to an outgroup[125]
- reducing intergroup uncertainty, anxiety and perceived threat[126] (see **4.3 Feelings of threat**)
- increasing empathy and perspective-taking[127]
- improving cognitive flexibility and hierarchical thinking.[128]

A 2015 Australian study using virtual contact combined with dual identity categorisation found indications that negative emotions were reduced first, in turn helping reduce intergroup bias (see **Case study 6.1**).[129]

Direct contact and virtual contact interventions provide the opportunity for intergroup personal relationships and friendships to form, and this may in turn encourage or provide further opportunities for intergroup contact.[130]

| **How effective the intervention is** | Contact interventions are one of the most highly researched and tested interventions. They are generally considered to be highly effective at reducing prejudice. During our consultations, multiple expert interviewees prioritised contact interventions, and several decades of research support the effectiveness of contact, particularly where they are structured and follow the 4 optimal conditions (see below).[131]

Research has found that:

- the prejudice-reduction effects of contact interventions tend to generalise to other members of the outgroup[132]
- contact interventions are still effective even among people with high levels of perceived threat[133] (see **4.3 Feelings of threat**) and in settings with a recent history of severe intergroup conflict.[134]

There is also evidence to indicate:

- both quality and quantity of contact have prejudice reduction effects, with quality contact (for example, more pleasant, natural, cooperative contact) being a greater predictor of lower prejudice[135]
- there is a causal link between intergroup contact between colleagues and reduced racial prejudice[136]
- there is a causal link between having had positive contact experiences and being less likely to believe in outgroup-directed conspiracy theories.[137]

Recent research also supports the role of less direct contact in reducing hate incidents. Notably, a 2023 Italian study found an association at a local level between increased consumption of cultural activity (for example, cinema, theatre, concerts, exhibitions) and a localised decrease in hate events over a 9-year period.[138] |

What factors increase its effectiveness	The original conception of contact interventions included 4 optimal conditions: ▪ equal status between the groups in the contact situation ▪ common goals ▪ intergroup cooperation ▪ support or sanction of a mutually recognised authority (for example, an employer, a school principal).[139] Follow-up research has supported the roles these 4 conditions play in increasing effectiveness[140], as well as the importance of: ▪ tailoring interventions to age groups and developmental stages[141] ▪ having a structured contact situation[142] ▪ setting up contact environments which allow participants to develop connections or friendships through meaningful and repeated contact.[143] Contact interventions are relatively flexible and can be combined with other prevention approaches (for example, categorisation[144]). A 2022 meta-analysis of prejudice-reduction studies in real-world settings found that contact interventions combined with awareness and/or perspective-taking are more effective than contact-only interventions – however, it should be noted that there were fewer studies with combination approaches than with contact-only approaches.[145]
What limitations and risks may be present	**Negative contact:** There is growing research into how negative contact experiences can lead to increased prejudice.[146] Current evidence indicates that negative contact interactions may have stronger effects (on increasing prejudice) than positive interactions (on decreasing prejudice); however, positive contact happens more frequently and is more likely during interventions.[147] Design and implementation should aim to reduce this risk by creating contact situations that encourage positive contact. **Cultural load and safety:** Depending on how a contact intervention is designed and run, there is a risk of cultural load, where participation could put an unequal burden on minority group member participants.[148] For example, the smaller number of minority group members in a school or workplace might mean that minority group members would have to invest more time and energy in contact situations (particularly direct and virtual contact). There may also be **safety risks**, where contact interventions may place people at risk of experiencing prejudice and hate from other participants. Design and implementation should aim to reduce these risks and provide support options for the wellbeing of minority group member participants. **'Ironic' side effect:** Although concern has been raised that intergroup contact might have an 'ironic' side effect of reducing support among minority members for social change, a 2022 meta-analysis found insufficient evidence to support this concern (and instead identified alternative explanations).[149]

Scalability: Direct and virtual contact interventions may face challenges around scalability. For example, in-person contact interventions can be hard to deliver at a larger scale given the time required of both organisers and participants, potential challenges in recruiting and getting buy-in from both majority group and minority group participants, and the physical constraints of having enough places or large enough places to facilitate them.[150] This is one reason that expert participants and evidence frequently highlight schools as a setting for contact interventions.

Charter rights: Public authorities that deliver contact interventions should consider how rights under the Charter, including freedom of expression and privacy, may be engaged and limited in these interventions.[151] For contact interventions involving children, public authorities should also consider whether the intervention engages the right under the Charter to the protection of children as is needed in their best interests.[152] If limitations are placed on rights, public authorities will need to consider whether those limitations are demonstrably justified.[153]

5.4 Perceived variability

How the intervention works

Perceived variability interventions are based on research linking perceived outgroup homogeneity bias – the tendency for people to perceive outgroups and their members as being essentially all the same – with the formation of stereotypes and prejudice.[154] The intervention aims to challenge this by demonstrating the diversity within a particular group, in turn reducing the tendency to form and apply generalisations and stereotypes.

Perceived variability is a relatively new intervention, first formally proposed in the early 2010s. We found 2 main approaches in the literature:

- **static posters** highlighting differences among members of an outgroup (for example, posters displaying photographs of several members of the same outgroup, their first names, ages, and a personality characteristic unique to each of them)[155]

- **short videos** highlighting differences among members of an outgroup (for example, videos where several members of the same overseas-born outgroup compare their experiences in Australia and overseas about everyday topics such as work–life balance, health and wellbeing).[156]

	Perceived variability interventions share similarities with counter-stereotypical or diverse representations of minority group members in the media. The main difference is that perceived variability interventions to date have presented multiple group members at the same point in time and highlighted their differences within the same umbrella characteristic (for example, in terms of personality traits or different perspectives on the same issue). General media representations may present one or multiple group members at once, and with different framings or contexts.
How effective the intervention is	There is insufficient research to draw conclusions on the effectiveness of perceived variability. Of the 2 studies we reviewed, there were mixed findings: ▪ early poster-based experiments in France were effective in reducing prejudice against Arab Muslim people[157] ▪ recent video-based attempts to extend and replicate the effect in Australia were not effective.[158] The authors of the Australian study proposed 3 possible reasons for the difference in effectiveness: the change in format, the different migration and societal contexts between countries, and the possibility of **false positives**[n] in the French study. Further research is needed to determine whether the original promising findings about perceived variability interventions can be replicated.
What factors increase its effectiveness	It is unclear from the evidence we reviewed what factors increase the effectiveness of perceived variability in preventing prejudice or vilification. As noted, prevention research continues to be conducted and published, and future research may find factors which increase the effectiveness of perceived variability.
What limitations and risks may be present	**Media-based interventions:** Broadly, media-based prejudice-reduction interventions have had mixed effectiveness, with some past campaigns unintentionally reinforcing stereotypes (see also **5.1 Awareness**).[159] This could have implications for media-based perceived variability interventions. **Unintended negative consequences:** There may be risks in designing representations intended to reflect the diversity of a particular community or identity (for example, inadvertently creating new stereotypes or generalisations, misrepresenting communities). Design and implementation should involve community consultation and testing with multiple audiences to avoid any unintended consequences.

---

n   False positives are when results indicate there was an effect when there was no effect. These can happen for a range of reasons (for example, poor research design and practices, inappropriate measures, low sample sizes, external factors interfering with what is being measured, mere chance). False positives are an ongoing concern in experimental research and highlight the importance of replicating studies to confirm whether the effects are real (see also **2.5 Caution in interpreting findings** and **3 Best practice principles**).

5.5 Perspective-taking

How the intervention works	Perspective-taking interventions aim to reduce prejudice by helping someone put themselves in the shoes of an outgroup member and find commonalities between their experiences and the outgroup member's.[160] The core idea is to encourage cognitive empathy (being able to identify someone else's mental state) and emotional empathy (being able to respond with the appropriate emotions).[161] Past research has tested various approaches to perspective-taking, including: ▪ writing a short essay from the perspective of an outgroup member ▪ physically simulating the experience of another (for example, climbing a staircase while wearing restrictive knee pads to simulate mobility restrictions) ▪ using virtual reality to make a person see themselves as a member of another group ▪ exchanging narratives in one-on-one conversations.[162] This report focuses on the last of these approaches, as recent research indicates it may be the most promising form of perspective-taking. In narrative exchanges, canvassers engage with prejudiced people in 10–15 minute conversations to non-judgementally exchange narratives and prompt a process of revising prejudicial attitudes.[163] The conversations may take different forms, but the core of the conversation is when the canvasser shares an experience of theirs (for example, a time they faced judgement or exclusion, a time someone showed them compassion when they needed it) and then asks the participant to share an experience of their own. This may be followed by explicitly encouraging reflections about an outgroup's similar experiences or about common misconceptions.
How effective the intervention is	Interpersonal perspective-taking conversations have emerged as a promising and scalable approach in the last decade. There are 2 main studies concerning racial or religious outgroups: ▪ A set of 3 randomised controlled trials tested one-on-one conversations with 6,869 people in 7 locations in the USA on attitudes towards unauthorised immigrants (as well as trans and gender diverse people).[164] It found that the intervention reduced prejudice for at least 4 months after the conversations. ▪ A randomised control trial tested interpersonal conversations on Islamophobia with 227 people in Melbourne.[165] Participants in the intervention group were aged between 43 and 94. It found that the intervention reduced prejudice for at least 12 weeks after the conversations.
What factors increase its effectiveness	More research is needed to determine what factors increase the effectiveness of interpersonal perspective-taking conversations (for example, whether effectiveness varies depending on the demographics of the canvassers). The US trials found that conversations that used only arguments were not effective, whereas conversations where narratives were exchanged were effective.[166]

What limitations and risks may be present	**Scepticism:** Some forms of perspective-taking may be received by the broader community with scepticism (for example, the use of virtual reality to see oneself as a member of another group).
	Cultural load: Interpersonal perspective-taking conversations involve the use of canvassers, some of whom may be members of the outgroup being discussed. There may be a risk of cultural load and burden on outgroup member canvassers. Design and implementation should aim to reduce this risk and provide support options for the wellbeing of minority group member participants.
	Charter rights: Public authorities that deliver perspective-taking interventions should consider how rights under the Charter, such as freedom of expression, may be engaged and limited in interventions.[167] If limitations are placed on rights, public authorities will need to consider whether those limitations are demonstrably justified.[168]

5.6 Other notable prevention approaches

Promising new directions in prevention approaches have emerged in recent years. In addition to the 5 main approaches explored above, other approaches were present in the literature we reviewed or were raised by experts or community members. While the 5 main approaches have been more extensively researched and evaluated, there is less evidence for these other approaches (or they were not entirely in the scope of our research questions), and more research and evaluation are needed. These other notable approaches include a mix of different mechanisms, contexts, cohorts and opportunities to intervene.

5.6.1 Early intervention for individuals at risk

There are opportunities for early identification, diversion and intervention of individuals at risk of hate conduct or radicalisation, regardless of age, using a strengths- and needs-based approach. These approaches aim to address the social issues that coincide with prejudice and hate.

Sustained disengagement from extremism and social reintegration requires a holistic approach across multiple domains:

- positive social relationships and connections
- coping with personal issues
- separation of identity from a cause or group
- shifts in ideology and behaviours towards non-violence and respect.[169]

Experts told us that deradicalisation and reintegration after radicalisation are long journeys that require funding and support to increase their chances of success. While diversion and deradicalisation are generally one-on-one processes, this is an important opportunity to disrupt hate groups and the influence of their individual members.

5.6.2 Early disruption of patterns of hate

There are also early-intervention opportunities at a larger scale through identification, monitoring and disruption of triggers, patterns and narratives. This can include the use of monitoring tools to proactively identify patterns that lead to or generate hate, including:

- likely trigger events
- emerging online vilification trends
- networks of perpetrators.
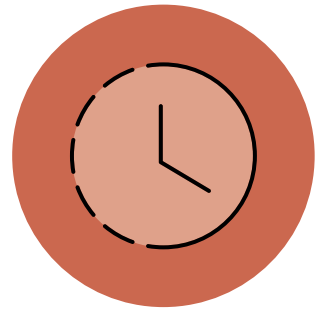
One expert spoke of the need for a local framework for early-response measures to prevent escalation and spread of hate. This could include a business-as-usual approach to campaigns, where they are ready to go and can responsively ramp up as required, similar to the way bushfire and public health campaigns are activated or increased at certain times. For example, campaigns could be activated in response to emerging trends or proactively in the lead-up to certain significant days or events known to be associated with increased hate.

Recent research supports the importance of identifying highly connected nodes in online hate networks and removing these nodes to disrupt the spread of online hate.[170] Given the reciprocal relationship between offline and online vilification, this may also be an opportunity to prevent offline behaviours.

5.6.3 School-based interventions

Schools are a commonly noted context for prejudice-reduction programs (see **Case study 6.1, Case study 6.3** and the recent Speak Out Against Racism (SOAR) program pilot[171]). These interventions and approaches are aimed at children and young people, when prejudice may be easier to prevent or address (see **4.2 When prejudice starts**). The importance and potential of school-based interventions were also recognised in 2 of the Inquiry's recommendations:

- **recommendation 5**, for primary school programs to strengthen respect, diversity and cohesion[172]
- **recommendation 6**, for clearer understanding among educators and school leaders on preventing and responding to hate conduct in schools.[173]
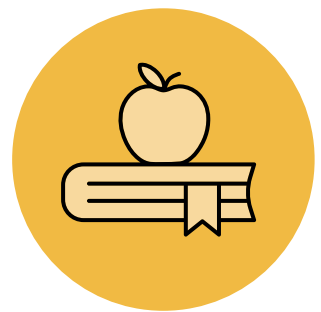
Reaching a critical mass or tipping point of the population to achieve a substantial shift in norms, attitudes and behaviours is challenging for interventions. Schools are an opportunity to scale up an intervention to reach a large proportion of the population over a sustained period in order to encourage and embed prosocial and anti-prejudice norms, attitudes and behaviours. School-based interventions may combine approaches such as awareness, contact, social norms and perspective-taking.

There is evidence supporting some key factors in school-based interventions:

▪ integrating awareness (see **5.1 Awareness**) and contact (see **5.3 Contact**) approaches to allow students to make personal connections across groups in a supportive environment and in ways that are meaningful and relevant to their lives

▪ supporting the development of teachers' own personal and professional intercultural capabilities

▪ supporting classroom-level teaching practice with whole-of-school leadership and administrative support.[174]

Schools can also be sites for early identification and intervention of children and young people at risk (see **5.6.1 Early intervention for individuals at risk**). One expert spoke about the potential to provide supports for parents and teachers on how to identify whether a child is being exposed to or demonstrating prejudice or hate, and next steps for parents, teachers and the child.

5.6.4 Social norms

Some interventions aim to shift, influence and enforce social norms and behaviours (see **Case study 6.3**). These social norm–based approaches can include:

▪ challenging existing norms

▪ modelling and reinforcing anti-racist and anti-hate norms

▪ educating and encouraging people to be role models and influencers

▪ representing and highlighting non-prejudicial behaviour in the media

▪ disrupting or preventing normalisation of and desensitisation to prejudice and vilification.[175]

Social norms can also be modelled and reinforced through bystander intervention and counterspeech (see **5.6.5 Prevention through response**).

5.6.5 Prevention through response

While this report focuses on approaches to preventing vilification in the first place, there is also a role for responses to vilification. Although responses to incidents do not deeply engage with the underlying drivers of prejudice or vilification, they may reinforce anti-hate norms (see **5.6.4 Social norms**) or prevent future behaviour or escalation. However, it is difficult to confidently establish a causal link between someone being challenged about something they have said and their subsequent attitudes or behaviours.
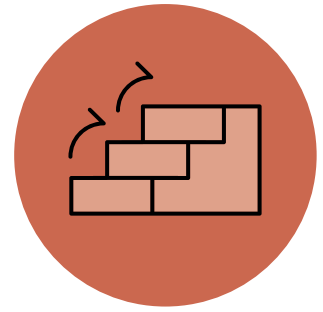
Bystander intervention

Bystander intervention is when an individual who witnesses or is present at a vilification act intervenes in the situation. This type of intervention aims to shift the burden of responding to vilification away from targets of vilification.[176] There is mixed evidence on its potential effectiveness as a prevention strategy:

- A 2011 literature review on bystander anti-racism found signs that bystander intervention could educate perpetrators, challenge perceptions of what it is acceptable to say, and shift social norms and attitudes.[177]

- A 2024 scoping review of bystander interventions in the workplace found indications that bystander action resulted in positive behaviour change from perpetrators.[178]

However, past studies have reported risks of backlash, perceived overreaction, negative personal consequences and negative professional consequences for those who confront prejudicial behaviours, particularly when a minority member confronts the behaviour.[179]

Counterspeech

Counterspeech is when an individual, organisation or bot[o] responds to vilification. Like bystander intervention, counterspeech aims to shift the burden of response away from targets and to challenge prejudicial views or vilification. Research on counterspeech has increased and diversified in recent years, including emerging research into the potential for automated generation of counterspeech using large language models (a form of generative artificial intelligence).[180]

A 2022 study of over 130,000 political conversations on German Twitter (now known as X) over 4 years found signs that organised and coordinated counterspeech was associated with reduction in the use of hate speech and a more neutral and balanced tone of conversation (see **Case study 6.2.1**).[181]

A 2021 real-world experiment with 1,350 Twitter users found indications that empathy-based counterspeech had some effect on perpetrators' later tweets and likelihood to delete a hateful tweet. No similar effect was found for counterspeech that used humour or warned of consequences (see **Case study 6.2.2**).[182]

---

o   A bot is a software application that runs automated tasks online, particularly tasks which are too repetitive or large scale for humans to do.

Researchers have noted the possibility that counterspeech's effectiveness in reducing hate behaviours may be contextual.[183] A perpetrator's behaviour may be displaced or dispersed from the original context where the counterspeech was delivered to a different context (for example, if they encounter counterspeech on one social media platform, they may stop vilifying on that platform and instead vilify on another).[184] At a larger scale, evidence suggests that the 2021 suspension of Parler, a social media platform associated with conspiracy theorists and far-right extremists, led to users migrating to other similar platforms but did not lead to a significant decrease in overall user activity on 'fringe' social media.[185]

5.6.6 Counternarratives

Counternarrative interventions aim to help people rewrite the biased narratives they believe. These are a kind of awareness intervention (see **5.1 Awareness**) that work by revealing and addressing the underlying 'logic' of those biased narratives. Although counternarratives are generally used in a violent radicalisation context,[186] they are notable for prejudice and vilification because of the links between narratives and biased stereotypes.

Some counternarrative approaches can shift how someone perceives threats, favours members of their ingroup and is hostile to outgroup members. A 2020 systematic review found that the most effective counternarrative approach involved challenging a dominant narrative with representations or examples that went against outgroup stereotypes. Other promising approaches involved the presentation of both the dominant narrative and an alternative, and an individual creating their own counterarguments.[187]

5.6.7 Inoculation against misinformation

Misinformation and disinformation are difficult to challenge and correct. There is continuing and increasing research into how to counter misinformation's influence effectively.[188] Evidence to date supports inoculation as a way to reduce misinformation's influence by:

- warning people that they will be exposed to challenging content – this allows people to be ready to think more critically about the information they are seeing and have their guard up against misinformation
- 'prebunking' through pre-emptively correcting a false claim with counterarguments or explaining manipulation techniques or questionable sources.[189]

A 2010 meta-analysis of inoculation demonstrated its effectiveness as an intervention and found that resistance to misinformation could generalise to new arguments encountered by participants.[190] Research into inoculation supports its effectiveness, including in countering extremist propaganda.[191] Most notably, a 2021 randomised trial in the context of Islamophobic and radical Islamist disinformation showed that inoculation reduced participants' likelihood to share content, feelings of anger and agreement with the content.[192]

5.6.8 System- and industry-level interventions

While this report focuses on prevention approaches aimed at shifting individual attitudes and behaviours, there are interventions at the system or industry levels that are worth noting.

Strengthened national regulation

Community members, experts and researchers have called for stronger regulation, standards and enforcement from national authorities to prevent and respond to prejudice and vilification in mass media and online media.[193] Strengthened national regulation of social media and mass media is not explored further here, as this report focuses on prevention approaches that could generally be implemented at a state level.

Content moderation

Content moderation is when an online platform or community representative, moderator, administrator or bot responds to hate content by removing, editing or filtering the content. While content moderation and counterspeech are related, they are different: counterspeech involves others *responding to* offensive content, whereas content moderation involves others *deleting or editing* the content (see **5.6.5 Prevention through response**). In some instances, content moderation can lead to action against and consequences for the user who created the content.

There is little research establishing the effectiveness of content moderation in preventing future hate content. Initial research into the effectiveness of hate speech content moderation on Twitter indicates that it might not deter perpetrators from continuing to engage on the platform.[194] There is also initial evidence that explaining a hate speech content moderation decision can help with the perceived trustworthiness of the decision.[195]

Further research is needed to establish the effectiveness of content moderation. This is particularly true considering recent research linking exposure to hate speech with desensitisation, dehumanisation and increased prejudice, potentially supporting the importance of removing content (see **4.6 Exposure to hate**). Public authorities that moderate content (for example, on government social media pages) should consider how this may engage and limit the right under the Charter to freedom of expression, such as measures to prevent harm.[196] Under the Charter, rights can be limited where this is demonstrably justified, such as with measures to prevent harm.[197]

Legal and regulatory deterrents

The existence of laws and regulations – including civil penalties, criminal offences, context-specific policies and regulatory schemes – may deter potential perpetrators from vilifying. However, it is difficult to confidently establish a causal link between the existence of a deterrent (like a state law) and individual behaviour when there are many other possible influences present.[198]

There is mixed evidence on deterrents as a way to *prevent* vilification. A 2021 study found evidence of a causal relationship between Germany's online hate speech laws and decreases in the volume of hate content and the intensity of religion- or migration-related hate content.[199] Using different data sources and methods, a 2015 study of Australian hate speech laws heard from community participants that the laws had not anecdotally led to decreased hate content, although the participants still supported the existence of the laws.[200]

This does not detract from the importance of legal and regulatory deterrents in terms of signalling what is acceptable by governments and communities,[201] enabling enforcement and providing remedies,[202] encouraging civility[203] and helping people feel safer in certain environments or on certain platforms.[204] Legal and regulatory deterrents can also encourage or require other deterrents (for example, by requiring organisations or online platforms to develop or strengthen their own deterrents).

Concern about a chilling effect

While concern is sometimes raised that anti-vilification laws might have a chilling effect on legitimate non-vilifying speech, no evidence appears to support this.[205] Notably, Germany's online hate speech law appeared to reduce the hate intensity only in religion- and migration-related tweets and not in other tweets – suggesting the law was not affecting other content from the same users.[206] Anti-vilification laws may engage and limit the right under the Charter to freedom of expression.[207] Under the Charter, lawmakers need to ensure that where rights are limited, those limitations are demonstrably justified.[208]

Enforcement of deterrents

Societal deterrents can be enabled through the (actual or possible) presence or oversight of others who may disapprove, intervene or enforce law, policy or regulation. In offline settings, this could be friends, strangers or law enforcement officers. Online, this could be moderators or other users, including on platforms that require non-anonymity. A 2013 study into young adults and online aggression found a relationship between a young adult's confidence in not getting caught and whether they engaged in aggressive online behaviour.[209]

The Commission has heard a range of views from First Peoples, racialised and marginalised religious Victorians about the presence of law enforcement officers. While some do feel safer with their presence, others have reported feeling uncomfortable and unsafe around law enforcement officers for a range of historical and current reasons. This does not detract from community support for the existence and enforcement of vilification laws.
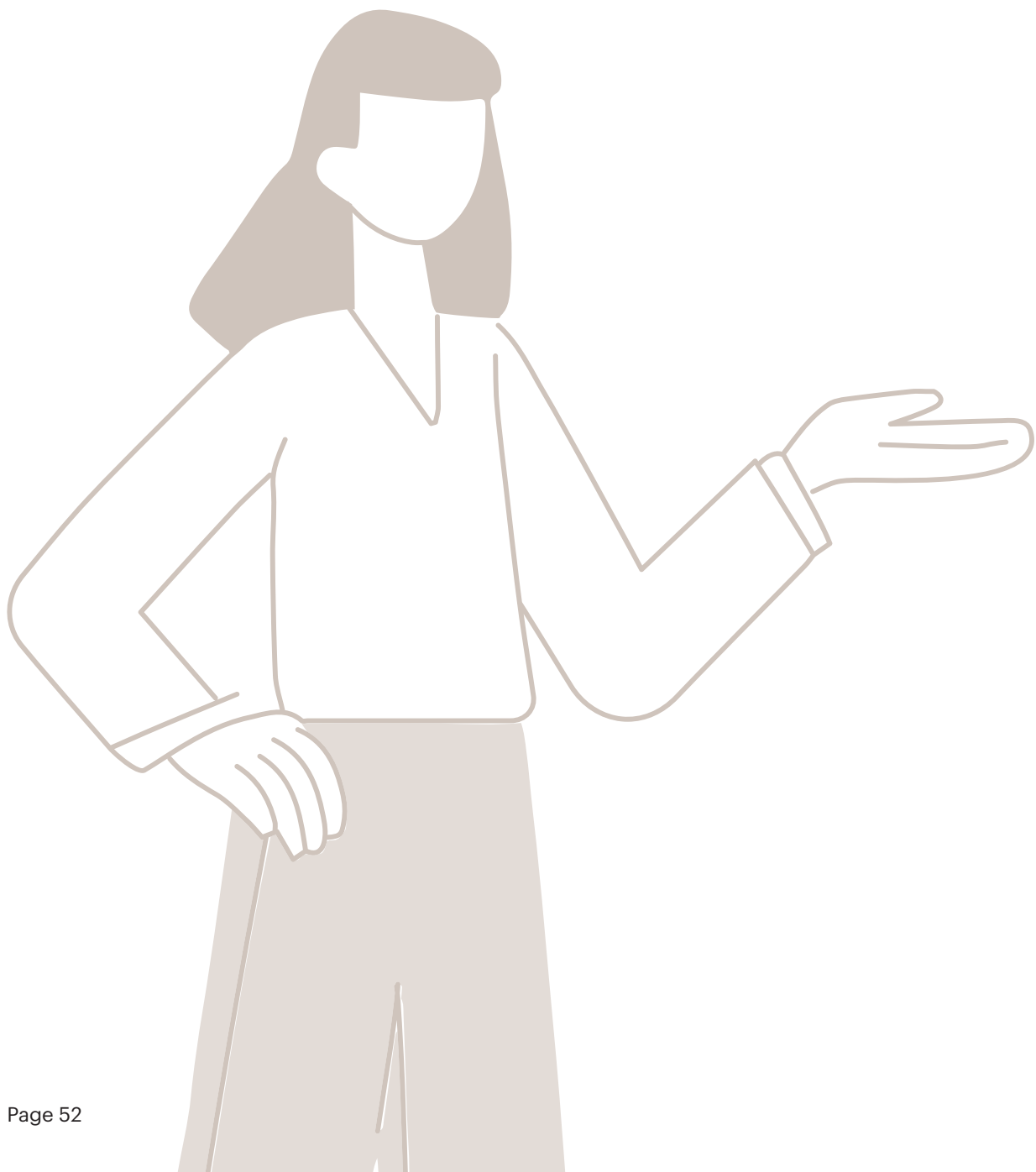
6. Case studies

To illustrate how prevention approaches can work, this section presents a series of case studies from real-world settings. The case studies are drawn from published peer-reviewed academic papers to demonstrate some of the prevention approaches discussed in the report, either by themselves or in combination. We have aimed to select examples which are accessible, practical and potentially applicable to the Victorian context.

6.1 Using online contact to reduce prejudice

COUNTRY	YEAR	PARTICIPANTS	PREVENTION APPROACHES USED
Australia	2015	Muslim and Christian high school students	5.1 Awareness 5.2 Categorisation 5.3 Contact 5.6.3 School-based interventions

A 2015 Australian real-world experiment investigated how the internet can improve intergroup relations.[210] Researchers looked at how emotions expressed during online intergroup contact might predict a reduction in intergroup bias between 102 Muslim and 102 Christian high school students aged around 12 years. Students in participating classes were allocated to one of 2 conditions:

- a Dual Identity Electronic Contact (DIEC) program integrating interfaith information (see **5.1 Awareness**), dual identity categorisation (see **5.2 Categorisation**) and intergroup contact between Christian and Muslim students (see **5.3 Contact**)

- a control group program involving only within-faith information and ingroup contact (that is, Christian students with Christian students, Muslim students with Muslim students).

During the 8-week DIEC program, Muslim and Christian students formed online groups and communicated with each other via a specially designed internet chat room. They discussed a social issue that was equally important to both groups: an environmentally sustainable Australia.

The DIEC program was designed to meet the 4 optimal conditions for successful intergroup contact (see **5.3 Contact**) and encourage participants to adopt dual identity recategorisation (see **5.2 Categorisation**). This involved simultaneously activating both their existing subgroup identity (in this case, as Christian or Muslim) and a new superordinate identity as pro-environment Australians. This meant they incorporated their respective religious beliefs and practices while working together towards the common goal of environmental sustainability.

Students completed a measure of intergroup bias 6 months before the intervention and then again at 2 weeks, 6 months and 12 months after the program. Online text exchanges between students were also analysed by examining the extent to which they used words reflecting 6 emotion categories – emotionality in general, positive emotions, negative emotions, anger, anxiety and sadness – negating the need for self-reported data.

The study found that compared to the control group, the DIEC program:

- increased participants' emotion-sharing in general (affect words) and their use of positive emotion words
- decreased their use of anger and sadness
- marginally decreased their use of negative emotion words.

The DIEC program participants also had significantly lower intergroup bias across all post-intervention time periods. This was positively associated with reduced anger and sadness.

This study is part of a body of research that demonstrates the value of intergroup contact helping improve intergroup relations. The study also provides some evidence that internet-based interactions can, like in-person interactions, play a role in prejudice reduction in adolescents that is sustained for up to 12 months.

6.2 Counterspeech on social media

6.2.1 Organised counterspeech

PLATFORM	YEAR	PARTICIPANTS	PREVENTION APPROACHES USED
Twitter (Germany)	2015–2018	An organised hate group An organised counter-hate group Twitter users	5.6.5 Prevention through response

A 4-year longitudinal study from 2015 to 2018 analysed 131,366 political conversations on German Twitter.[211] These conversations involved 2 organised groups: a far-right hate group formed in 2016 (Reconquista Germanica, or RG) and a counter-hate group formed in early 2018 (Reconquista Internet, or RI). Drawing from theories on bullying and social norms, the authors hypothesised that RI's organised counterspeech (see **5.6.5 Prevention through response**) would be more effective in reducing prevalence of hate speech than would counterspeech from non-organised individuals.

The analysis found that when RI became active, there was a more balanced presence of both hate and counterspeech. RI's organised counterspeech was more effective than individual efforts in steering the tone of conversations. This was mostly by:

- providing more (and organised) support to counterspeech tweets
- guiding neutral discussions towards counterspeech.

The authors suggested that, as in traditional bullying settings, providing support to peers can motivate people to oppose hate speech and defend those targeted by it. They concluded that organised counterspeech can positively alter online discussions and may help curb hateful speech.

6.2.2 Use of empathy

PLATFORM	YEAR	PARTICIPANTS	PREVENTION APPROACHES USED
Twitter	2020–2021	Twitter users	5.6.5 Prevention through response

A real-world experiment conducted from 2020 to 2021 among 1,350 Twitter users tested different types of counterspeech.[212] It classified a sample of tweets as containing xenophobic (or racist) hate speech and randomly assigned them to either a control group or one of 3 response tweet interventions:

- counterspeech intended to induce empathy and humanise the person targeted (for example, 'For African Americans, it really hurts to see people use language like this')
- counterspeech that warned of potential consequences (for example, 'Hey, remember that your friends and family can see this tweet too')
- humour- or meme-based counterspeech (for example, a humorous meme of an animal captioned with 'It's time to stop tweeting').

The response tweets were publicly visible, generated by a human-controlled bot account and sent within 24 hours of the original tweet.

Analysis indicated that empathy-based counterspeech was the only intervention to have any effect. When the authors converted the data to tweets, it was reported that, in the 4-week experimental period, users receiving the empathy treatment sent (on average) 1.3 fewer xenophobic tweets and were 8.4% more likely to delete the original tweet.

6.2.3 Trusted community organisations

PLATFORM	YEAR	PARTICIPANTS	PREVENTION APPROACHES USED
Twitter	2015–2016	Jewish community organisations Antisemitic Twitter users Twitter users	5.6.5 Prevention through response

A UK cross-sectional study on antisemitism on Twitter collected tweets over a one-year period between 2015 and 2016.[213] It tested several hypotheses, including whether users who were perceived as trustworthy and capable (in this case, Jewish community organisations) were associated with:

- more retweets of their content
- a longer 'survival' period of retweets (the length of time between the first and last retweet of each piece of their content within the study's data collection period).

The study's findings indicated that there were significantly more retweets and longer survival periods for content from Jewish organisations than from antisemitic users. The authors suggested that this supports:

- the ability of positive content to challenge hate speech – as there is a higher volume of positive content, and it spreads further and survives longer

- the value of trusted community organisations on social media in increasing collective efficacy against hate

- the role of trusted community organisations in countering vilification.

6.3 Preventing hate speech among adolescents

COUNTRY	YEAR	PARTICIPANTS	PREVENTION APPROACHES USED
Germany	2022	High school students	5.1 Awareness
			5.5 Perspective-taking
			5.6.3 School-based interventions
			5.6.4 Social norms
			5.6.5 Prevention through response
			5.6.7 Inoculation against misinformation

A 2022 study evaluated the short-term effects of a hate speech prevention program known as HateLess Together against Hatred.[214] It looked at adolescents' empathy towards targets of hate speech, their self-efficacy in dealing with hate speech incidents and their engagement in counterspeech.

The HateLess program aimed to prevent hate speech perpetration and victimisation among adolescents and to equip them with skills to stand up against hate speech among their peers. The program comprised 5 modules delivered by teachers (not external trainers) over one week, combining activities on individual, classroom and school levels. The modules broadly covered:

- understanding hate speech and the differences between hate speech and the rights to freedom of opinion and expression (see **5.1 Awareness**)

- causes and motivators of hate speech, including social norms and potential digital amplifiers of hate speech, such as misinformation (see **5.1 Awareness** and **5.6.7 Inoculation against misinformation**)

- empathy-building through methods that included perspective-taking (see **5.5 Perspective-taking**)

- effective individual coping strategies to deal with hate speech victimisation

- how to intervene in hate speech and support targeted classmates, including expressing criticism without hurting the feelings of others (see **5.6.4 Social norms** and **5.6.5 Prevention through response**)

- sharing what was learned with others in the school community.

To examine the effects of the HateLess program, the study recruited 820 German school children aged 12 to 16 years. Students were assigned to either an intervention group, which participated in the HateLess program, or a control group, which did not receive any program.

The study used self-report measures at 2 time periods: 1 week before the program and 1 month after. Students self-reported their empathy for victims of hate speech, self-efficacy towards intervening in hate speech and engagement in counterspeech.

Results indicated that in the short term, compared with adolescents in the control group, adolescents who received the HateLess intervention self-reported:

▪ feeling higher levels of empathy for victims of hate speech

▪ feeling more self-efficacious towards intervening

▪ countering hate speech more often.

The authors concluded that the program was an effective and cost-efficient way to strengthen adolescents' use of counterspeech and indirectly train them in key civics skills. They noted that follow-up data would be required to see if the changes persisted over the longer term and that changes could be verified from sources other than self-reports. They also called for further research to identify the characteristics of non-responders in order to improve future programs.

7. Where to from here

Effective prevention strategies combine an understanding of the evidence on drivers and prevention with best practice principles. This section explores opportunities for this in practice in Victoria.

There is no known formula of drivers that would produce or prevent prejudice or vilification. However, there are key drivers that increase their likelihood:

- influences of family, friends and society – through opinions, behaviours and norms (see **4.2 When prejudice starts**)
- feelings of anxiety or perceived threat from other groups (see **4.3 Feelings of threat**)
- negative stereotypes and narratives, particularly when spread by and through the media following trigger events (see **4.4 Media narratives and trigger events**)
- normalisation of and desensitisation to prejudice, discrimination and hate (see **4.5 When prejudice becomes hate**)
- repeated exposure to hate, which can lead to desensitisation, dehumanisation, limited ability to empathise, and potential perpetration (see **4.6 Exposure to hate**)
- the roles of online spaces and communities – some of which bond over hate, use platforms to recruit and grow, and spread and normalise hate (see **4.7 Online hate communities**).

These drivers are important considerations when designing prevention strategies. Each prevention approach explored in this report targets different drivers in different ways. The evidence supports several promising and important approaches to preventing prejudice and vilification, in particular:

- intergroup contact (see **5.3 Contact**)
- dual identity recategorisation (see **5.2 Categorisation**)
- perspective-taking through conversation (see **5.5 Perspective-taking**)
- empathy-based counterspeech (see **5.6.5 Prevention through response**)
- inoculation against misinformation (see **5.6.7 Inoculation against misinformation**).

There are opportunities to bring these approaches together into prevention strategies in Victoria, using best practice principles to give them a greater chance of success and lasting impact (see **3 Best practice principles**):

1. identifying evidence-based change mechanisms and outcomes
2. tailoring strategies to the context they will be delivered in
3. considering scalability when selecting approaches
4. engaging with the evidence to learn from past lessons and limitations
5. using community-centred and -led design approaches where possible
6. integrating a behavioural insights lens
7. setting up rigorous outcome-focused monitoring and evaluation early
8. trialling strategies to test for intended, unintended and undesired outcomes
9. sharing knowledge and lessons learned to help build publicly available knowledge.

Several promising prevention approaches outlined in this report could have roles to play in Victoria's prevention strategies across different settings. These opportunities would require careful design, pilot testing, monitoring and evaluation that were context-specific, locale-specific and behaviourally informed. They would also require consideration of risks and limitations (for example, Charter rights, cultural load, unintended negative consequences).

The following table presents some example opportunities that could be refined through a full design and testing process.

Setting	Potential opportunity	Relevant report sections
Educational institutions, including schools and universities	An ongoing statewide multiple-approach program that could integrate: ▪ an awareness-based education component that also encouraged anti-prejudice norms and behaviours, including empathy-based responses to hate ▪ dual identity recategorisation through intergroup contact exercises or workshops ▪ perspective-taking narrative exchanges through interactive activities ▪ techniques that reduced the spread and impact of misinformation. The possibility of a template version of this program could be explored where, following pilot testing, it could be adapted to the local community, educational institution or and classroom context. Self-assessment tools could complement the template program and this report to allow individual educational institutions to develop tailored action plans.	*Prevention approaches* 5.1 Awareness 5.2 Categorisation 5.3 Contact 5.5 Perspective-taking 5.6.3 School-based interventions 5.6.4 Social norms 5.6.5 Prevention through response 5.6.7 Inoculation against misinformation *Case studies* 6.1 Using online contact to reduce prejudice 6.2.2 Use of empathy 6.3 Preventing hate speech among adolescents
Media campaigns	An ongoing statewide campaign across multiple channels that could: ▪ responsibly increase awareness of prejudice and hate (by avoiding pitfalls of past media campaigns) ▪ promote counter-stereotypical narratives ▪ encourage anti-prejudice norms and behaviours ▪ be rapidly and responsively ramped up or targeted to specific regions or contexts as required ▪ contribute to inoculation against misinformation. The possibility of a template version of this campaign could be explored where, following pilot testing, it could be adapted to particular communities, regions or trigger events.	*Prevention approaches* 5.1 Awareness 5.6.2 Early disruption of patterns of hate 5.6.4 Social norms 5.6.5 Prevention through response 5.6.6 Counternarratives 5.6.7 Inoculation against misinformation

Setting	Potential opportunity	Relevant report sections
Various, including community centres, events, schools and universities, and workplaces	A conversation-based program that could integrate: • perspective-taking narrative exchanges • indirect intergroup contact elements • counternarrative-style techniques • inoculation against misinformation.	*Prevention approaches* 5.3 Contact 5.5 Perspective-taking 5.6.6 Counternarratives 5.6.7 Inoculation against misinformation
Various, including media, online spaces, and schools and universities	A large-scale initiative to reduce exposure to (and the impact of) hate and harmful narratives through: • inoculation and counternarrative techniques adapted for larger scales • coordinated national approaches to regulation and content moderation • supports and services equipped to identify individuals at risk and address underlying needs • supports and services that provide engaging and attractive alternative spaces to hate communities.	*Prevention approaches* 5.6.1 Early intervention for individuals at risk 5.6.6 Counternarratives 5.6.7 Inoculation against misinformation 5.6.8 System- and industry-level interventions

Appendix: Additional literature

The following items of literature were analysed and have contributed to the Commission's understanding of drivers and prevention, but are not directly cited in the report.

Blaya, Catherine, 'Cyberhate: A review and content analysis of intervention strategies' (2019) 45 *Aggression and Violent Behavior* 163.

Bobba, Beatrice and Elisabetta Crocetti, '"I feel you!": The role of empathic competences in reducing ethnic prejudice among adolescents' (2022) 51(10) *Journal of Youth and Adolescence* 1970.

Cheng, Jennifer E, '"It's not a race, it's a religion": Denial of anti-Muslim racism in online discourses' (2022) 11(10) *Social Sciences* 467.

Cikara, Mina, Vasiliki Fouka and Marco Tabellini, 'Hate crime towards minoritized groups increases as they increase in sized-based rank' (2022) 6(11) *Nature Human Behaviour* 1537.

Costello, Matthew et al, 'COVID-19 and Sinophobia: Detecting warning signs of radicalization on Twitter and Reddit' (2023) 26(7) *Cyberpsychology, Behavior, and Social Networking* 546.

Czopp, Alexander M, Margo J Monteith and Aimee Y Mark, 'Standing up for a change: Reducing bias through interpersonal confrontation' (2006) 90(5) *Journal of Personality and Social Psychology* 784.

Dentice, Dianne, '"So much for Darwin" – an analysis of Stormfront discussions on race' (2019) 15(1) *Journal of Hate Studies* 133.

Garland, Joshua et al, 'Countering hate on social media: Large scale classification of hate and counter speech' (No 2006.01974, arXiv, 5 June 2020) <**http://arxiv.org/abs/2006.01974**>.

Hammond, Kate and Nesreen Bottriell, *Stand out, speak up: Racism and Islamophobia in Victorian schools* (The Australian Muslim Women's Centre for Human Rights, 2023).

Hawdon, James et al, 'The perpetuation of online hate: A criminological analysis of factors associated with participating in an online attack' (2019) 15(1) *Journal of Hate Studies* 157.

Hill, Miriam E and Martha Augoustinos, 'Stereotype change and prejudice reduction: Short- and long-term evaluation of a cross-cultural awareness programme' (2001) 11(4) *Journal of Community & Applied Social Psychology* 243.

Kauff, Mathias et al, 'How do we get people into contact? Predictors of intergroup contact and drivers of contact seeking' (2021) 77(1) *Journal of Social Issues* 38.

Lambe, Jennifer et al, 'Predictors of anti-Asian xenophobia during COVID-19: Towards developing targeted intervention campaigns' (2021) 17(1) *Journal of Hate Studies* 146.

Leung, Candace and Richard Frank, 'Unity starts with U: A case study of a counter-hate campaign through the use of social media platforms' (2020) 16(1) *Journal of Hate Studies* <**https://account.jhs.press.gonzaga.edu/index.php/gl-j-jhs/article/view/146**>.

Lowry, Paul Benjamin et al, 'Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model' (2016) 27(4) *Information Systems Research* 962.

MacNaughton, Glenda and Karina Davis, 'Beyond "othering": Rethinking approaches to teaching young Anglo-Australian children about Indigenous Australians' (2001) 2(1) *Contemporary Issues in Early Childhood* 83.

Mansouri, Fethi and Matteo Vergani, 'Intercultural contact, knowledge of Islam, and prejudice against Muslims in Australia' (2018) 66 *International Journal of Intercultural Relations* 85.

Matthes, Jörg and Desirée Schmuck, 'The effects of anti-immigrant right-wing populist ads on implicit and explicit attitudes: A moderated mediation model' (2015) 44(4) *Communication Research* 556.

McWaters, Samuel C and Russell Hawkins, 'The imagined contact hypothesis: Prejudice towards asylum seekers in Australia' (2018) 3(4) *International Journal of Innovation, Creativity and Change* 106.

Miklikowska, Marta, 'Empathy trumps prejudice: The longitudinal relation between empathy and anti-immigrant attitudes in adolescence' (2018) 54(4) *Developmental Psychology* 703.

Munden, Hanna Paalgard, 'Cognitive distortions in men who have exited white supremacist groups' [2023] *Terrorism and Political Violence* 1.

Nagin, Daniel S, 'Deterrence: A review of the evidence by a criminologist for economists' (2013) 5(1) *Annual Review of Economics* 83.

Obermaier, Magdalena and Desirée Schmuck, 'Youths as targets: Factors of online hate speech victimization among adolescents and young adults', ed Jessica Vitak (2022) 27(4) *Journal of Computer-Mediated Communication* zmac012.

Obermaier, Magdalena, Desirée Schmuck and Muniba Saleem, 'I'll be there for you? Effects of Islamophobic online hate speech and counter speech on Muslim in-group bystanders' intention to intervene' (2023) 25(9) *New Media & Society* 2339.

Paluck, Elizabeth Levy and Donald P Green, 'Prejudice reduction: What works? A review and assessment of research and practice' (2009) 60(1) *Annual Review of Psychology* 339.

Papcunová, Jana et al, 'Perception of hate speech by the public and experts: Insights into predictors of the perceived hate speech towards migrants' (2023) 26(7) *Cyberpsychology, Behavior, and Social Networking* 489.

Pedersen, Anne and Fiona Kate Barlow, 'Theory to social action: A university-based strategy targeting prejudice against Aboriginal Australians' (2008) 43(3) *Australian Psychologist* 148.

Pehrson, Samuel, Rupert Brown and Hanna Zagefka, 'When does national identification lead to the rejection of immigrants? Cross-sectional and longitudinal evidence for the role of essentialist in-group definitions' (2009) 48(1) *British Journal of Social Psychology* 61.

Pieterse, Tanya, Vanessa Stratford and Juan A Nel, 'Relationship between symbolic violence and overt violence in hate incidents in South Africa' (2018) 16(2) *African Safety Promotion: A Journal of Injury and Violence Prevention* 31.

Poole, Elizabeth, Eva Haifa Giraud and Ed De Quincey, 'Tactical interventions in online hate speech: The case of #stopIslam' (2021) 23(6) *New Media & Society* 1415.

Pratt, Travis C and Jillian J Turanovic, 'Celerity and deterrence' in Daniel S Nagin, Francis T Cullen and Cheryl Lero Jonson (eds), *Deterrence, choice, and crime, Volume 23: Contemporary perspectives* (Taylor & Francis, 2018) 187.

Priest, Naomi et al, '"You are not born being racist, are you?" Discussing racism with primary aged-children' (2016) 19(4) *Race Ethnicity and Education* 808.

Roxell, Lena, 'Hate, threats, and violence. A register study of persons suspected of hate crime' (2011) 12(2) *Journal of Scandinavian Studies in Criminology and Crime Prevention* 198.

Savelkoul, M et al, 'Anti-Muslim attitudes in the Netherlands: Tests of contradictory hypotheses derived from ethnic competition theory and intergroup contact theory' (2011) 27(6) *European Sociological Review* 741.

Simmons, Cécile et al, *Reciprocal dynamics between Australia's political fringes on Twitter* (Centre for Resilient and Inclusive Societies, 19 October 2021) <**https://www.crisconsortium.org/research-reports/political-fringes-twitter**>.

Turoy-Smith, Katrine M, Robert Kane and Anne Pedersen, 'The willingness of a society to act on behalf of Indigenous Australians and refugees: The role of contact, intergroup anxiety, prejudice, and support for legislative change' (2013) 43(Issue Supplement S2) *Journal of Applied Social Psychology* E179.

Vidgen, Bertram, 'Tweeting Islamophobia: Islamophobic hate speech amongst followers of UK political parties on Twitter' (PhD thesis, University of Oxford, 2019) <**https://ora.ox.ac.uk/objects/uuid:3c32d29d-e2e4-4913-abf8-2a28886f55a7**>.

Wachs, Sebastian et al, 'Associations between witnessing and perpetrating online hate speech among adolescents: Testing moderation effects of moral disengagement and empathy' (2022) 12(6) *Psychology of Violence* 371.

Wachs, Sebastian et al, 'The online hate speech cycle of violence: Moderating effects of moral disengagement and empathy in the victim-to-perpetrator relationship' (2022) 25(4) *Cyberpsychology, Behavior, and Social Networking* 223.

Wachs, Sebastian et al, 'Associations between classroom climate, empathy, self-efficacy, and countering hate speech among adolescents: A multilevel mediation analysis' (2023) 38(5–6) *Journal of Interpersonal Violence* 5067.

Wickes, Rebecca et al, *Understanding the context of racial and cultural exclusivism: A study of Melbourne neighbourhoods* (Monash University, April 2020) <**https://doi.org/10.26180/5e6ee4fa77a39**>.

Williams, Matthew L et al, 'The effect of the Brexit vote on the variation in race and religious hate crimes in England, Wales, Scotland and Northern Ireland' (2023) 63(4) *The British Journal of Criminology* 1003.

Windisch, Steven et al, 'Online interventions for reducing hate speech and cyberhate: A systematic review' (2022) 18(2) *Campbell Systematic Reviews* e1243.

Notes

1   *Racial and Religious Tolerance Act 2001* (Vic) Preamble.

2   Naomi Priest et al, *Racism, racial discrimination and child and youth health: A rapid evidence synthesis* (Victorian Health Promotion Foundation, 2021) <**https://doi.org/10.37309/2021.MW1027**>; Yin Paradies, 'Racism and Indigenous Health' in *Oxford Research Encyclopedia of Global Public Health* (Oxford University Press, 2018) <**https://oxfordre.com/publichealth/display/10.1093/acrefore/9780190632366.001.0001/acrefore-9780190632366-e-86**>; Office of the eSafety Commissioner, *Online hate speech: Findings from Australia, New Zealand and Europe* (Office of the eSafety Commissioner, 2020) <**https://www.esafety.gov.au/research/online-hate-speech**>; Derya Iner, Ron Mason and Chloe Smith, *Islamophobia in Australia IV (2014–2021)* (Islamophobia Register Australia, 2023) <**https://islamophobia.com.au/publications/islamophobia-in-australia-reports/**>; Jewish Community Council of Victoria and Community Security Group Victoria, *Victorian antisemitism report 30 June 2023* (Community Security Group Victoria, 30 June 2023) <**https://www.csgvic.com.au/antisemitismreport**>; Mohamed Hadi, *Racism at the University of Melbourne report, edition 3* (UMSU Inc, 2023); Matthew Williams, *The science of hate: How prejudice becomes hate and what we can do to stop it* (Faber & Faber, 2021); Colin Roberts et al, *Understanding who commits hate crime and why they do it* (Welsh Government Social Research, 2013); Susann Wiedlitzka et al, 'Hate in word and deed: The temporal association between online and offline Islamophobia' (2023) 39(1) *Journal of Quantitative Criminology* 75; Kathryn Benier et al, *'Don't drag me into this': Growing up South Sudanese in Victoria after the 2016 Moomba 'riot'* (Centre for Multicultural Youth, 2018) <**https://www.cmy.net.au/resource/dont-drag-me-into-this-growing-up-south-sudanese-in-victoria-after-the-2016-moomba-riot/**>; Diego A Díaz-Faes and Noemí Pereda, 'Is there such a thing as a hate crime paradigm? An integrative review of bias-motivated violent victimization and offending, its effects and underlying mechanisms' (2022) 23(3) *Trauma, Violence, & Abuse* 938; Kathryn Benier, Rebecca Wickes and Claire Moran, '"African gangs" in Australia: Perceptions of race and crime in urban neighbourhoods' (2021) 54(2) *Journal of Criminology* 220; Matteo Vergani et al, 'When and how does anti-semitism occur? The different trigger mechanisms associated with different types of criminal and non-criminal hate incidents' (2022) 43(9) *Deviant Behavior* 1135; David Tittensor, Gerhard Hoffstaedter and Adam Possamai, 'Hyper-securitisation and belonging: Understanding the plight of young Muslims in Melbourne, Australia' (2020) 40(3) *Journal of Muslim Minority Affairs* 480.

3   Damon Centola et al, 'Experimental evidence for tipping points in social convention' (2018) 360(6393) *Science* 1116; Manjana Milkoreit, 'Social tipping points everywhere? Patterns and risks of overuse' (2023) 14(2) *WIREs Climate Change* e813.

4   See, for example, Erin R Watson and Charles R Collins, 'Putting the system in systemic racism: A systems thinking approach to advancing equity' (2023) 71(3–4) *American Journal of Community Psychology* 274.

5   Gordon Willard Allport, *The nature of prejudice* (Addison-Wesley Publishing Company, 1954).

6   Legislative Assembly Legal and Social Issues Committee, *Inquiry into anti-vilification protections* (Parliament of Victoria, 3 March 2021) 65 <**https://new.parliament.vic.gov.au/4a287a/contentas-sets/56e922dff39041edb5d49ad3d6eaa808/inquiry_into_anti-vilification_protections_002.pdf**>.

7   Matteo Vergani et al, *Defining and identifying hate motives: Bias indicators for the Australian context* (Centre for Resilient and Inclusive Societies, October 2022) <**https://doi.org/10.56311/POZS1016**>.

8   Vincent Petit, *The behavioural drivers model: A conceptual framework for social and behaviour change programming* (UNICEF, 2019) <**https://www.unicef.org/mena/reports/behavioural-drivers-model**>.

9   Wing Hsieh, Rebecca Wickes and Nicholas Faulkner, 'What matters for the scalability of prejudice reduction programs and interventions? A Delphi study' (2022) 10(1) *BMC Psychology* 107, 5.

10   Jennifer Schweppe and Barbara Perry, 'A continuum of hate: Delimiting the field of hate studies' (2022) 77(5) *Crime, Law and Social Change* 503; Matteo Vergani and Rouven Link, 'A conceptual framework to map responses to hate crime, hate incidents and hate speech: The case of Australia' [2021] *Social Policy and Society* 1.

11   *Racial and Religious Tolerance Act 2001* (n 1) ss 7–8, 24–25.

12   Schweppe and Perry (n 10).

13 Mario Peucker, Tom Clark and Holly Claridge, 'Mapping the journey of (non-) reporting in response to racism: A change-oriented approach to reporting barriers, motives and support needs' [2023] *Journal of Intercultural Studies* <**https://doi.org/10.1080/07256868.2023.2296026**>; Matteo Vergani and Carolina Navarro, *Barriers to reporting hate crime and hate incidents in Victoria: A mixed-methods study* (Centre for Resilient and Inclusive Societies, 30 June 2020); Matteo Vergani and Carolina Navarro, 'Hate crime reporting: The relationship between types of barriers and perceived severity' (2023) 29(1) *European Journal on Criminal Policy and Research* 111.

14 Caroline Erentzen and Regina Schuller, 'Exploring the dark figure of hate: Experiences with police bias and the under-reporting of hate crime' (2020) 62(2) *Canadian Journal of Criminology & Criminal Justice* 64.

15 Hsieh, Wickes and Faulkner (n 9).

16 Wing Hsieh, Nicholas Faulkner and Rebecca Wickes, 'What reduces prejudice in the real world? A meta-analysis of prejudice reduction field experiments' (2022) 61(3) *British Journal of Social Psychology* 689.

17 Ibid.

18 Wing Hsieh, Nicholas Faulkner and Rebecca Wickes, 'Perceived variability as a video-media prejudice reduction intervention' (2022) 44(2) *Basic and Applied Social Psychology* 66.

19 Elizabeth Levy Paluck et al, 'Prejudice reduction: progress and challenges' (2021) 72(1) *Annual Review of Psychology* 533, 535.

20 Maureen McBride, *What works to reduce prejudice and discrimination? A review of the evidence* (The Scottish Government, 2015); Mark Sutton et al, *Getting the Message across: Using media to reduce racial prejudice and discrimination* (Communities and Local Government Publications, December 2007); Jehonathan Ben, David Kelly and Yin Paradies, 'Contemporary anti-racism: A review of effective practice' in John Solomos (ed), *Routledge International Handbook of Contemporary Racisms* (Routledge, 2020) 205 <**https://www.taylorfrancis.com/books/9781351047319/chapters/10.4324/9781351047326-16**>.

21 Hsieh, Faulkner and Wickes (n 16).

22 Sutton et al (n 20).

23 McBride (n 20); Priest et al (n 2); Sutton et al (n 20); Ben, Kelly and Paradies (n 20).

24 Sutton et al (n 20); Ben, Kelly and Paradies (n 20).

25 Petit (n 8).

26 See, for example, Williams (n 2).

27 McBride (n 20); Williams (n 2).

28 Chang Won Jung, 'Role of informal social control in predicting racist hate speech on online platforms: Collective efficacy and the theory of planned behavior' (2023) 26(7) *Cyberpsychology, Behavior, and Social Networking* 507.

29 Ana-Maria Bliuc et al, 'Online networks of racial hate: A systematic review of 10 years of research on cyber-racism' (2018) 87 *Computers in Human Behavior* 75; Elisabetta Crocetti et al, 'Development of prejudice against immigrants and ethnic minorities in adolescence: A systematic review with meta-analysis of longitudinal studies' (2021) 60 *Developmental Review* 100959.

30 Priest et al (n 2).

31 Tobias Raabe and Andreas Beelmann, 'Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta-analysis of age differences' (2011) 82(6) *Child Development* 1715; Williams (n 2).

32 Ioana Zagrean et al, 'The family transmission of ethnic prejudice: A systematic review of research articles with adolescents' (2022) 11(6) *Social Sciences* 236.

33 McBride (n 20); Roberts et al (n 2).

34 Matteo Vergani, Fethi Mansouri and Liliana Orellana, 'Terrorism concern and persistence of negative attitudes towards Islam and Muslims' (2022) 32(6) *Journal of Community & Applied Social Psychology* 1029.

35 Safiyya Khan and Anne Pedersen, 'Black African immigrants to Australia: Prejudice and the function of attitudes' (2010) 4(2) *Journal of Pacific Rim Psychology* 116; Anne Pedersen et al, 'Attitudes toward Indigenous Australians and asylum seekers: The role of false beliefs and other social-psychological variables' (2005) 40(3) *Australian Psychologist* 170.

36 Ryan Perry, Chris G Sibley and John Duckitt, 'Dangerous and competitive worldviews: A meta-analysis of their associations with social dominance orientation and right-wing authoritarianism' (2013) 47(1) *Journal of Research in Personality* 116; Melisa Castellanos et al, 'Direct and indirect effects of social dominance orientation on hate speech perpetration via empathy and moral disengagement among adolescents: A multilevel mediation model' [2023] *Aggressive Behavior* <**https://onlinelibrary.wiley.com/doi/abs/10.1002/ab.22100**>; Williams (n 2).

37 Khan and Pedersen (n 35h).

38 Priest et al (n 2).

39   Kevin Dunn et al, 'Countering Islamophobia in the Victorian population' [2021] (3) *CRIS Issues Papers* <**https://doi.org/10.56311/CEMO4002**>.

40   Walter G Stephan and White Stephan Cookie, 'An integrated threat theory of prejudice' in Stuart Oskamp (ed), *Reducing Prejudice and Discrimination* (Psychology Press, 2000) 23 <**https://doi.org/10.4324/9781410605634**>; Jasper Van Assche et al, 'Intergroup contact is reliably associated with reduced prejudice, even in the face of group threat and discrimination' (2023) 78(6)*American Psychologist 761*; Robert Schweitzer et al, 'Attitudes towards refugees: The dark side of prejudice in Australia' (2005) 57(3) *Australian Journal of Psychology* 170.

41   Helen Forbes-Mewett and Rebecca Wickes, 'The neighbourhood context of crime against international students' (2018) 54(4) *Journal of Sociology* 609; Benier et al (n 2); Vergani, Mansouri and Orellana (n 34).

42   Stephan and Cookie (n 40); Mario Peucker, *Media reporting on far-right extremism in Australia: Between strategic silence and harmful amplification* (Centre for Resilient and Inclusive Societies, May 2023); Schweitzer et al (n 40); Vergani, Mansouri and Orellana (n 34).

43   Schweitzer et al (n 40); Vergani, Mansouri and Orellana (n 34).

44   D Mark Anderson, Benjamin Crost and Daniel I Rees, 'Do economic downturns fuel racial animus?' (2020) 175 *Journal of Economic Behavior & Organization* 9.

45   Daria Denti and Alessandra Faggian, 'Where do angry birds tweet? Income inequality and online hate in Italy' (2021) 14(3) *Cambridge Journal of Regions, Economy and Society* 483.

46   Benier et al (n 2); Zoe Leviston, Justine Dandy and Jolanda Jetten, '"They're discriminated against, but so are we": White Australian born perceptions of ingroup and immigrant discrimination over time are not zero sum' (2021) 60(1) *British Journal of Social Psychology* 146; Matthew L Williams and Pete Burnap, 'Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data' (2016) 56(2) *The British Journal of Criminology* 211.

47   Vergani, Mansouri and Orellana (n 34).

48   Bliuc et al, 'Online networks of racial hate: A systematic review of 10 years of research on cyber-racism' (n 29); Purpose, *Online hate speech in Australia: The role of news media and pathways for change – Part 1: The dynamics of hate speech online* (Portable, 25 September 2023) <**https://www.purpose.com/australian-news-media-and-online-hate-speech/**>; Roberts et al (n 2); Williams (n 2); Wiedlitzka et al (n 2); Vergani, Mansouri and Orellana (n 34); Enqi Weng and Fethi Mansouri, '"Swamped by Muslims" and facing an "African gang" problem: Racialized and religious media representations in Australia' (2021) 35(3) *Continuum* 468; Matthew Costello et al, 'COVID-19: A pandemic of anti-Asian cyberhate' (2021) 17(1) *Journal of Hate Studies* 108; Williams and Burnap (n 45); Julia R DeCook and Mi Hyun Yoon, 'Kung flu and roof Koreans: Asian/Americans as the hated other and proxies of hating in the white imaginary' (2021) 17(1) *Journal of Hate Studies* 119.

49   Roberts et al (n 2); Sutton et al (n 20); Brian Griffiths and Anne Pedersen, 'Prejudice and the function of attitudes relating to Muslim Australians and Indigenous Australians' (2009) 61(4) *Australian Journal of Psychology* 228; Purpose (n 48).

50   Purpose (n 48).

51   Weng and Mansouri (n 48); Roberts et al (n 2).

52   Williams and Burnap (n 46); Christian Von Sikorski, Jörg Matthes and Desirée Schmuck, 'The Islamic State in the news: Journalistic differentiation of Islamist terrorism from Islam, terror news proximity, and Islamophobic attitudes' (2021) 48(2) *Communication Research* 203.

53   Purpose (n 48).

54   Roberts et al (n 2); Benier et al (n 2); Von Sikorski, Matthes and Schmuck (n 52).

55   Díaz-Faes and Pereda (n 2); Vergani, Mansouri and Orellana (n 34); Matteo Vergani et al, 'Hate speech in a Telegram conspiracy channel during the first year of the COVID-19 pandemic' (2022) 8(4) *Social Media + Society* <**https://doi.org/10.1177/20563051221138758**>; Williams and Burnap (n 45).

56   Díaz-Faes and Pereda (n 2); Williams (n 2); Benier, Wickes and Moran (n 2); Benier et al (n 2); Sara Cheikh Husain and Fethi Mansouri, 'Australian Muslims' visibility: The politics of oppression and recognition' (2023) 14(1) *Religions* 93; Vergani, Mansouri and Orellana (n 34); Vergani et al, 'When and how does anti-semitism occur? The different trigger mechanisms associated with different types of criminal and non-criminal hate incidents' (n 55); Imran Awan and Irene Zempi, '"I will blow your face off" – Virtual and physical world anti-Muslim hate crime' (2015) 57(2) *The British Journal of Criminology* 362.

57   Vergani et al (n 2).

58 Vergani, Mansouri and Orellana (n 34).

59 Weng and Mansouri (n 48).

60 Von Sikorski, Matthes and Schmuck (n 52).

61 Williams and Burnap (n 46).

62 Weng and Mansouri (n 48); Griffiths and Pedersen (n 49).

63 Matteo Vergani, Thierno Diallo and Kerry O'Brien, 'Measuring the potential for hateful behaviours: Development and validation of the Hate Behaviours Scale (HBS)' [2023] *Terrorism and Political Violence* <**https://doi.org/10.1080/09546553.2023.2283565**>.

64 Benier, Wickes and Moran (n 2).

65 Kevin Dunn et al, *Understanding varied attitudes towards Muslims* (PDF, Western Sydney University, 2021) <**https://doi.org/10.26183/bkkk-0195**>; Schweitzer et al (n 40).

66 Khan and Pedersen (n 35); Griffiths and Pedersen (n 49).

67 Schweitzer et al (n 40); Stephan and Cookie (n 40).

68 Leonie Huddy et al, 'Anger, anxiety, and selective exposure to terrorist violence' (2021) 65(10) *Journal of Conflict Resolution* 1764.

69 Roberts et al (n 2); Díaz-Faes and Pereda (n 2).

70 Griffiths and Pedersen (n 49).

71 Alanna Kamp et al, 'Asian Australians' experiences of online racism during the COVID-19 pandemic' (2022) 11(5) *Social Sciences* 227; Costello et al (n 48).

72 Ana-Maria Bliuc et al, 'The effects of local socio-political events on group cohesion in online far-right communities' (2020) 15(3) *PLoS ONE* e0230302.

73 Griffiths and Pedersen (n 49); Kamp et al (n 71); Costello et al (n 48); Bliuc et al, 'The effects of local socio-political events on group cohesion in online far-right communities' (n 72).

74 See, for example, Ian Anderson et al, 'Racism and the 2023 Australian constitutional referendum' (2023) 402(10411) *The Lancet* 1400.

75 See, for example, Díaz-Faes and Pereda (n 2); Williams (n 2); Roberts et al (n 2).

76 EF Haghish et al, 'Mental health, well-being, and adolescent extremism: A machine learning study on risk and protective factors' (2023) 51(11) *Research on Child and Adolescent Psychopathology* 1699.

77 Williams (n 2).

78 Bettina Rottweiler, Caitlin Clemmow and Paul Gill, 'A common psychology of male violence? Assessing the effects of misogyny on intentions to engage in violent extremism, interpersonal violence and support for violence against women' [2024] *Terrorism and Political Violence* <**https://www.tandfonline.com/doi/full/10.1080/09546553.2023.2292723**>.

79 Nicholas Faulkner and Ana-Maria Bliuc, '"It's okay to be racist": Moral disengagement in online discussions of racist incidents in Australia' (2016) 39(14) *Ethnic and Racial Studies* 2545; Jade Hutchinson et al, 'Online and offline coordination in Australia's far-right: A study of True Blue Crew' (2022) 11(9) *Social Sciences* 421; Matheus Schmitz et al, 'Do users adopt extremist beliefs from exposure to hate Subreddits?' (2024) 14(1) *Social Network Analysis and Mining* 22.

80 M Kim, M Ellithorpe and SA Burt, 'Anonymity and its role in digital aggression: A systematic review' (2023) 72 *Aggression and Violent Behavior* 101856.

81 Wiktor Soral, Michał Bilewicz and Mikołaj Winiewski, 'Exposure to hate speech increases prejudice through desensitization' (2018) 44(2) *Aggressive Behavior* 136; Agnieszka Pluta et al, 'Exposure to hate speech deteriorates neurocognitive mechanisms of the ability to understand others' pain' (2023) 13(1) *Scientific Reports* 4127.

82 Justin Cheng et al, 'Anyone can become a troll: Causes of trolling behavior in online discussions' in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (ACM, 2017) 1217 <**https://dl.acm.org/doi/10.1145/2998181.2998213**>.

83 Julia Kansok-Dusche et al, 'A systematic review on hate speech among children and adolescents: Definitions, prevalence, and overlap with related phenomena' (2023) 24(4) *Trauma, Violence, & Abuse* 2598.

84 Karsten Müller and Carlo Schwarz, 'Fanning the flames of hate: Social media and hate crime' (2021) 19(4) *Journal of the European Economic Association* 2131; Costello et al (n 48); Matthew L Williams et al, 'Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime' (2019) 60 *The British Journal of Criminology* 93.

85 Vergani, Mansouri and Orellana (n 34); Weng and Mansouri (n 48); Williams et al (n 84); All Together Now, *Social commentary and racism in 2019* (All Together Now, November 2019) <**https://alltogethernow.org.au/social-commentary-and-racism-in-2019/**>.
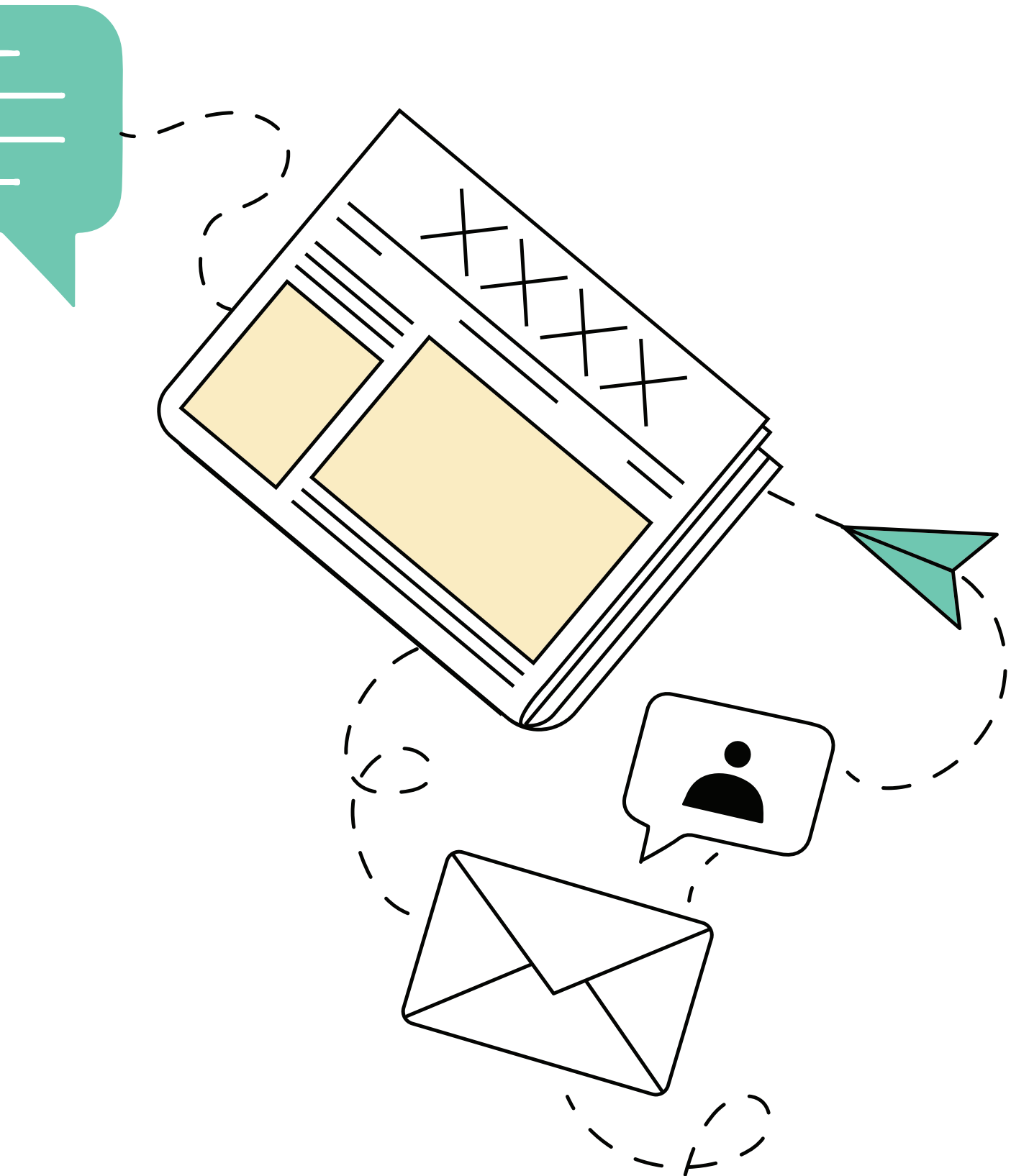
86 Karsten Müller and Carlo Schwarz, 'From hashtag to hate crime: Twitter and antiminority sentiment' (2023) 15(3) *American Economic Journal: Applied Economics* 270.

87 Costello et al (n 48).

88 Bliuc et al, 'Online networks of racial hate: A systematic review of 10 years of research on cyber-racism' (n 29).

89 Minwoong Chung, Seyoung Lee and Heejo Keum, 'Seeing is behaving? The role of social norms in the relationship between witnessing and perpetrating online hate' [2023] *Communication Reports* 1; Michael Hameleers, Toni van der Meer and Rens Vliegenthart, 'Civilized truths, hateful lies? Incivility and hate speech in false information – evidence from fact-checked statements in the US' (2022) 25(11) *Information, Communication & Society* 1596; Sebastian Wachs et al, 'Playing by the rules? An investigation of the relationship between social norms and adolescents' hate speech perpetration in schools' (2022) 37(21–22) *Journal of Interpersonal Violence* NP21143; Wiedlitzka et al (n 2); Sebastian Wachs et al, 'Online correlates of cyberhate involvement among young people from ten European countries: An application of the routine activity and problem behaviour theory' (2021) 123 *Computers in Human Behavior* 106872; Matthew Costello, Salvatore J Restifo and James Hawdon, 'Viewing anti-immigrant hate online: An application of routine activity and social structure-social learning theory' (2021) 124 *Computers in Human Behavior* 106927; Tracey J Hayes, 'Trump's digital rhetoric of hate: The use of enthymemes in creating division' (2021) 17(1) *Journal of Hate Studies* 14; Müller and Schwarz (n 84); Schmitz et al (n 79).

90 Bliuc et al, 'Online networks of racial hate: A systematic review of 10 years of research on cyber-racism' (n 29); Hutchinson et al (n 79); M Rohangis Mohseni, 'Motives of online hate speech: Results from a quota sample online survey' (2023) 26(7) *Cyberpsychology, Behavior, and Social Networking* 499; Ana-Maria Bliuc et al, 'Collective identity changes in far-right online communities: The role of offline intergroup conflict' (2019) 21(8) *New Media & Society* 1770.

91 Bliuc et al, 'The effects of local socio-political events on group cohesion in online far-right communities' (n 72).

92 Müller and Schwarz (n 84); Müller and Schwarz (n 86); Schmitz et al (n 79).

93 Wachs et al, 'Playing by the rules? An investigation of the relationship between social norms and adolescents' hate speech perpetration in schools' (n 89); Wiedlitzka et al (n 2); Wachs et al, 'Online correlates of cyberhate involvement among young people from ten European countries: An application of the routine activity and problem behaviour theory' (n 89).

94 See, for example, McBride (n 20); Sutton et al (n 20).

95 Hsieh, Faulkner and Wickes (n 16); McBride (n 20); Sutton et al (n 20).

96 Legislative Assembly Legal and Social Issues Committee (n 6) 98.

97 Hsieh, Faulkner and Wickes (n 16); Katerina Bezrukova et al, 'A meta-analytical integration of over 40 years of research on diversity training evaluation' (2016) 142(11) *Psychological Bulletin* 1227; Ben, Kelly and Paradies (n 20).

98 McBride (n 20); Sutton et al (n 20); Bezrukova et al (n 97).

99 Bezrukova et al (n 97).

100 Jessica Walton, Naomi Priest and Yin Paradies, 'Identifying and developing effective approaches to foster intercultural understanding in schools' (2013) 24(3) *Intercultural Education* 181.

101 Bezrukova et al (n 97).

102 Ben, Kelly and Paradies (n 20).

103 Ibid.

104 Ayoub Bouguettaya et al, 'I won't listen if I think we're losing our way: How right-wing authoritarianism affects the response to different anti-prejudice messages', ed Shrisha Rao (2023) 18(1) *PLoS ONE* e0280557.

105 Sutton et al (n 20) 32 (Figure 4.1).

106 Ibid 44 (Table 5.1).

107 McBride (n 20); Sutton et al (n 20); Walton, Priest and Paradies (n 100).

108 Bezrukova et al (n 97).

109 Vergani, Mansouri and Orellana (n 34).

110 Stephan Lewandowsky et al, 'Misinformation and its correction: Continued influence and successful debiasing' (2012) 13(3) *Psychological Science in the Public Interest* 106.

111 Andreas Beelmann and Kim Sarah Heinemann, 'Preventing prejudice and improving intergroup attitudes: A meta-analysis of child and adolescent training programs' (2014) 35(1) *Journal of Applied Developmental Psychology* 10; Lindsey Cameron et al, 'Changing children's intergroup attitudes toward refugees: Testing different models of extended contact' (2006) 77(5) *Child Development* 1208.

112 Fiona A White et al, 'Emotion expression and intergroup bias reduction between Muslims and Christians: Long-term internet contact' (2015) 53 *Computers in Human Behavior* 435; Cameron et al (n 111); Hsieh, Faulkner and Wickes (n 16).

113 Jessica Walton et al, 'Talking culture? Egalitarianism, color-blindness and racism in Australian elementary schools' (2014) 39 *Teaching and Teacher Education* 112; Ben, Kelly and Paradies (n 20).

114 Hsieh, Faulkner and Wickes (n 16).

115 Beelmann and Heinemann (n 111).

116 White et al (n 112).

117 Cameron et al (n 111).

118 Beelmann and Heinemann (n 111).

119 *Charter of Human Rights and Responsibilities Act 2006* (Vic) s 17(2).

120 See Allport (n 5).

121 Hsieh, Faulkner and Wickes (n 16); Thomas F Pettigrew and Linda R Tropp, 'A meta-analytic test of intergroup contact theory' (2006) 90(5) *Journal of Personality and Social Psychology* 751; Gordon Hodson, 'Interracial prison contact: The pros for (socially dominant) cons' (2008) 47(2) *British Journal of Social Psychology* 325.

122 Gunnar Lemmer and Ulrich Wagner, 'Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions' (2015) 45(2) *European Journal of Social Psychology* 152; Shelly Zhou et al, 'The extended contact hypothesis: A meta-analysis on 20 years of research' (2019) 23(2) *Personality and Social Psychology Review* 132.

123 Zhou et al (n 122); Daria Denti, Alessandro Crociata and Alessandra Faggian, 'Knocking on hell's door: Dismantling hate with cultural consumption' (2023) 47(2) *Journal of Cultural Economics* 303.

124 Birte Siem, Lisa Neymeyer and Anette Rohmann, 'Entertainment education as a means to reduce anti-Muslim prejudice – for whom does it work best?: An extended replication of Murrar and Brauer (2018)' (2021) 52(1) *Social Psychology* 51.

125 Pettigrew and Tropp (n 121); Lemmer and Wagner (n 122).

126 Hsieh, Faulkner and Wickes (n 16); Pettigrew and Tropp (n 121); Nils Karl Reimer et al, 'Building social cohesion through intergroup contact: Evaluation of a large-scale intervention to improve intergroup relations among adolescents' (2021) 50(6) *Journal of Youth and Adolescence* 1049; Cameron et al (n 111).

127 Hsieh, Faulkner and Wickes (n 16); Daniel Jolley, Charles R Seger and Rose Meleady, 'More than a prejudice reduction effect: Positive intergroup contact reduces conspiracy theory beliefs' (2023) 53(6) *European Journal of Social Psychology* 1262; Reimer et al (n 126).

128 Jolley, Seger and Meleady (n 127).

129 White et al (n 112).

130 Beelmann and Heinemann (n 111); Zhou et al (n 122).

131 Hsieh, Faulkner and Wickes (n 16); Lemmer and Wagner (n 122); Pettigrew and Tropp (n 121).

132 Lemmer and Wagner (n 122).

133 Van Assche et al (n 40).

134 Lemmer and Wagner (n 122).

135 Jolley, Seger and Meleady (n 127); Jens Binder et al, 'Does contact reduce prejudice or does prejudice reduce contact? A longitudinal test of the contact hypothesis among majority and minority groups in three European countries' (2009) 96(4) *Journal of Personality and Social Psychology* 843.

136 Sean Darling-Hammond, Randy T Lee and Rodolfo Mendoza-Denton, 'Interracial contact at work: Does workplace diversity reduce bias?' (2021) 24(7) *Group Processes & Intergroup Relations* 1114; Williams (n 2).

137 Jolley, Seger and Meleady (n 127).

138 Denti, Crociata and Faggian (n 123).

139 Pettigrew and Tropp (n 121).

140 Ibid.

141 Cameron et al (n 111).

142 Pettigrew and Tropp (n 121).

143 Ben, Kelly and Paradies (n 20).

144 Cameron et al (n 111); White et al (n 112).

145 Hsieh, Faulkner and Wickes (n 16).

146 Fiona Kate Barlow et al, 'The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice' (2012) 38(12) *Personality and Social Psychology Bulletin* 1629.

147 Sarina J Schäfer et al, 'Does negative contact undermine attempts to improve intergroup relations? Deepening the understanding of negative contact and its consequences for intergroup contact research and interventions' (2021) 77(1) *Journal of Social Issues* 197.

148 Ihsan Yilmaz et al, 'Young Muslim Australians' experiences of intergroup contact and its implications for intercultural relations' (2020) 44 *Ethnic and Racial Studies* 2772.

149 Nils Karl Reimer and Nikhil Kumar Sengupta, 'Meta-analysis of the "ironic" effects of intergroup contact' (2023) 124(2) *Journal of Personality and Social Psychology* 362.

150 Hsieh, Wickes and Faulkner (n 9).

151 *Charter of Human Rights and Responsibilities Act 2006* (n 118) s 15.

152 Ibid s 17(2).

153 Ibid s 7(2).

154 Abdelatif Er-rafiy and Markus Brauer, 'Modifying perceived variability: Four laboratory and field experiments show the effectiveness of a ready-to-be-used prejudice intervention' (2013) 43(4) *Journal of Applied Social Psychology* 840; Hsieh, Faulkner and Wickes (n 18).

155 Er-rafiy and Brauer (n 154).

156 Hsieh, Faulkner and Wickes (n 18).

157 Er-rafiy and Brauer (n 154).

158 Hsieh, Faulkner and Wickes (n 18).

159 Sutton et al (n 20).

160 Hsieh, Faulkner and Wickes (n 16).

161 Kathryn Benier et al, 'Reducing Islamophobia through conversation: A randomized control trial' [2024] *Social Psychology Quarterly* 01902725231217246.

162 Paluck et al (n 19).

163 Joshua L Kalla and David E Broockman, 'Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments' (2020) 114(2) *American Political Science Review* 410; Benier et al (n 161).

164 Kalla and Broockman (n 163).

165 Benier et al (n 161).

166 Kalla and Broockman (n 163).

167 *Charter of Human Rights and Responsibilities Act 2006* (n 118) s 15.

168 Ibid s 7(2).

169 Kate Barrelle, 'Pro-integration: Disengagement from and life after extremism' (2015) 7(2) *Behavioral Sciences of Terrorism and Political Aggression* 129.

170 Wafa Alorainy et al, 'Disrupting networks of hate: Characterising hateful networks and removing critical nodes' (2022) 12(1) *Social Network Analysis and Mining* 27.

171 Naomi Priest et al, *Evaluation of the Speak Out Against Racism (SOAR) program pilot* (No 2, ANU Centre for Social Research & Methods, 2020) <https://csrm.cass.anu.edu.au/research/publications/evaluation-speak-out-against-racism-soar-program-pilot>.

172 Legislative Assembly Legal and Social Issues Committee (n 6) 79.

173 Ibid 84.

174 Walton, Priest and Paradies (n 100); Priest et al (n 171).

175 Hsieh, Faulkner and Wickes (n 16).

176 Jacqueline K Nelson, Kevin M Dunn and Yin Paradies, 'Bystander anti-racism: A review of the literature' (2011) 11(1) *Analyses of Social Issues and Public Policy* 263.

177 Ibid.

178 Laura Jennings et al, 'Mapping bystander intervention to workplace inclusion: A scoping review' (2024) 34(2) *Human Resource Management Review* 101017.

179 Ben, Kelly and Paradies (n 20); Jennings et al (n 178); Alexander M Czopp and Margo J Monteith, 'Confronting prejudice (literally): Reactions to confrontations of racial and gender bias' (2003) 29(4) *Personality and Social Psychology Bulletin* 532.

180 Raj Ratn Pranesh, Ambesh Shekhar and Anish Kumar, 'Towards automatic online hate speech intervention generation using pretrained language model' (OpenReview, 2021) <https://openreview.net/forum?id=0y5zhDzIQxq>; Wanzheng Zhu and Suma Bhat, 'Generate, prune, select: A pipeline for counterspeech generation against online hate speech' (No 2106.01625, arXiv, 3 June 2021) <http://arxiv.org/abs/2106.01625>; Punyajoy Saha et al, 'CounterGeDi: A controllable approach to generate polite, detoxified and emotional counterspeech' (No 2205.04304, arXiv, 9 May 2022) <http://arxiv.org/abs/2205.04304>.

181 Joshua Garland et al, 'Impact and dynamics of hate and counter speech online' (2022) 11(1) *EPJ Data Science* 3.

182 Dominik Hangartner et al, 'Empathy-based counterspeech can reduce racist hate speech in a social media field experiment' (2021) 118(50) *Proceedings of the National Academy of Sciences* e2116310118.

183 Ibid; Sutton et al (n 20).

184 Sutton et al (n 20).

185 Manoel Horta Ribeiro et al, 'Deplatforming did not decrease Parler users' activity on fringe social media' (2023) 2(3) *PNAS Nexus* pgad035.

186 Sébastien Brouillette-Alarie et al, 'Systematic review on the outcomes of primary and secondary prevention programs in the field of violent radicalization' [2022] (30) *Journal for Deradicalization* 117.

187 Sarah L Carthy et al, 'Counter-narratives for the prevention of violent radicalisation: A systematic review of targeted interventions' (2020) 16(3) *Campbell Systematic Reviews* e1106.

188 Lewandowsky et al (n 110).

189 Stephan Lewandowsky and Sander van der Linden, 'Countering misinformation and fake news through inoculation and prebunking' (2021) 32 *European Review of Social Psychology*.

190 John A Banas and Stephen A Rains, 'A meta-analysis of research on inoculation theory' (2010) 77(3) *Communication Monographs* 281.

191 Lewandowsky and van der Linden (n 189); Kurt Braddock, 'Vaccinating against hate: using attitudinal inoculation to confer resistance to persuasion by extremist propaganda' (2022) 34(2) *Terrorism and Political Violence* 240.

192 Stephan Lewandowsky and Muhsin Yesilada, 'Inoculating against the spread of Islamophobic and radical-Islamist disinformation' (2021) 6(1) *Cognitive Research: Principles and Implications* 57.

193 All Together Now (n 85); Purpose (n 48).

194 Rafael Jiménez Durán, 'The economics of content moderation: Theory and experimental evidence from hate speech on Twitter' [2023] *SSRN Electronic Journal* <**https://www.ssrn.com/abstract=4044098**>.

195 Sai Wang and Ki Joon Kim, 'Content moderation on social media: Does it matter who and why moderates hate speech?' (2023) 26(7) *Cyberpsychology, Behavior, and Social Networking* 527.

196 *Charter of Human Rights and Responsibilities Act 2006* (n 119) s 15.

197 Ibid s 7(2).

198 Travis C Pratt et al, 'The empirical status of deterrence theory: A meta-analysis' in Francis T Cullen, John Paul Wright and Kristie R Blevins (eds), *Taking stock: The status of criminological theory* (Transaction Publishers, 2008) 367 <**https://doi.org/10.4324/9781315130620**>.

199 Raphaela Andres and Olga Slivko, 'Combating online hate speech: The impact of legislation on Twitter' [2021] *SSRN Electronic Journal* <**https://www.ssrn.com/abstract=4013662**>.

200 Katharine Gelber and Luke McNamara, 'The effects of civil hate speech laws: Lessons from Australia' (2015) 49(3) *Law & Society Review* 631.

201 Williams (n 2).

202 Andres and Slivko (n 199).

203 Gelber and McNamara (n 200).

204 Ina Weber et al, 'Features for hate? Using the Delphi method to explore digital determinants for online hate perpetration and possibilities for intervention' (2023) 26(7) *Cyberpsychology, Behavior, and Social Networking* 479.

205 Gelber and McNamara (n 200).

206 Andres and Slivko (n 199).

207 *Charter of Human Rights and Responsibilities Act 2006* (n 119) s 15.

208 Ibid s 7(2).

209 Michelle F Wright, 'The relationship between young adults' beliefs about anonymity and subsequent cyber aggression' (2013) 16(12) *Cyberpsychology, Behavior, and Social Networking* 858.

210 White et al (n 112).

211 Garland et al (n 181).

212 Hangartner et al (n 182).

213 Sefa Ozalp et al, 'Antisemitism on Twitter: Collective efficacy and the role of community organisations in challenging online hate speech' (2020) 6(2) *Social Media + Society* <**http://journals.sagepub.com/doi/10.1177/2056305120916850**>.

214 Sebastian Wachs et al, 'Effects of the prevention program "HateLess. Together against Hatred" on adolescents' empathy, self-efficacy, and countering hate speech' (2023) 52(6) *Journal of Youth and Adolescence* 1115.

**Victorian Equal Opportunity
& Human Rights Commission**