Canadian Journal of Philosophy (2024), 54: 7, 482–506
doi:10.1017/can.2025.10017

DISTINGUISHED LECTURE

A Harm Reduction Approach to Attitudinal Racism

Joseph Heath

Department of Philosophy, University of Toronto, Toronto, ON, Canada
Email: joseph.heath@utoronto.ca

Abstract
Although the harm-reduction approach to policy is most familiar from debates over public health and drug abuse, it provides a perfectly general framework for thinking about normative aspects of policy in non-ideal contexts. This paper seeks to apply a generalized harm reduction approach to the problem of attitudinal racism. Psychological research suggests that racism is unlikely to be completely eradicated, as a result of which a zero-tolerance approach risks becoming both counterproductive and overly punitive. Harm reduction recommends minimization of prevalence with respect to the primary phenomenon combined with attenuation of impact for the ineliminable portion.

Keywords: racism; harm reduction; public policy; anti-racism; authoritarianism; intolerance

Most human beings alive today inhabit an environment that is profoundly unnatural, in the sense that it differs radically from the one in which our species evolved. This poses innumerable challenges, as we seek to regulate both our thought and behavior in order to suppress or redirect psychological dispositions that, left unchecked, would lead us to dysfunctional or socially maladaptive responses. The presence of artificial light, ultra-processed high-calorie foods, refined drugs and stimulants, and even photographic or video representations, constitute environmental changes that threaten to evoke maladaptive behavioral responses, ranging from insomnia and obesity to addiction and anxiety (Stanovich, 2004). The same is true not just with respect to the physical environment, but the social environment as well. The ancestral environment was one in which human beings engaged in repeated interactions over time with a small number of well-known individuals, many of them close relatives. Modern societies, by contrast, feature frequent one-time interactions with individuals who are, and essentially remain, strangers. The social world that most of us inhabit is a great deal more chaotic, unpredictable, challenging, and diverse than anything encountered even by our recent ancestors.

It should hardly come as a surprise, therefore, to find that living in such an environment is psychologically taxing. In particular, it requires a great deal more effortful attention to navigate successfully, because it forces us constantly to overcome potentially deleterious intuitive reactions from our autonomous psychological systems (Heath, 2014; Stanovich, 2004, 136–138). Furthermore, because there is natural variation in the strength of these underlying psychological dispositions, not to mention differences in levels of achieved self-control, the challenge of navigating modern social environments is a great deal more taxing for some than for others. For example, retributive impulses in response to norm-violation are an intuitive psychological response, which

© The Author(s), 2025. Published by Cambridge University Press on behalf of The Canadian Journal of Philosophy, Inc. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

are felt more strongly by some people (Carlsmith & Darley, 2008, 213–216). As a result, the civil condition, in which the state reserves for itself the prerogative of punishing various forms of wrongdoing, imposes much greater burdens of "instinctual renunciation" on some than others. Many remain perpetually dissatisfied with the relatively mild and delayed forms of retribution meted out by bureaucratic systems of justice.

As I will attempt to show in this paper, a similar situation prevails with respect to toleration of pluralism and group differences. The ancestral social environment was both smaller in scale and a great deal more homogeneous than contemporary societies. Everyday conduct was governed by a set of well-known rules that applied, with few exceptions, to all persons. Everyone spoke the same language, ate the same food, shared the same habits of grooming and dress, and respected the same customs. As a result, individuals were seldom called upon to exercise what we now refer to as "tolerance," which involves refraining from responding punitively to those who appear to be flouting these norms, but are in fact just following a *different* set of norms. Modern societies require significant cultivation of this virtue, because they routinely impose on individuals the obligation to act cooperatively (or at least non-aggressively) toward individuals with whom they have very little in common, and no knowledge of past behavior. Again, as research on authoritarian personality traits has shown, respect for this obligation imposes much greater psychological and motivational burdens on some people than on others (Stenner, 2012).

Failures of toleration show up in many guises, the most conspicuous of which is racism, understood narrowly to refer to derogatory or antagonistic attitudes toward out-group members demarcated by observable physiological differences linked to ancestry (there is, in fact, a somewhat complex definitional debate about both racism and its motivating psychology, which I shall strive to clarify in the discussion below[1]). The important point is that the psychological complex underlying racist beliefs and behavior, although no doubt subject to socializing influences, remains in certain key respects persistent, and so must be actively monitored by individuals and in many cases suppressed. Again, this is likely to prove more demanding for some than others, based on both the strength of the underlying psychological dispositions and the robustness of the capacity for executive control. Unfortunately, a great deal of the policy response to societal racism has been based on the claim that racism is learned behavior, and so can be eliminated entirely (among those who have not been exposed) or unlearned (among those who have).[2] Because racism is also subject to strong moral disapproval, this has tended to favor a "zero tolerance" approach to the problem, which combines the policy objective of complete eradication with a strong punitive response to all instances of transgression.

My central objective in this paper will be to defend what I refer to, following Daniel Weinstock (2020) and Shannon Dea (2020), as a "harm reduction" policy approach, applied in this case to attitudinal racism. A central tenet of social identity theory is that racism is potentiated by a set of ubiquitous, enduring psychological dispositions (Tajfel & Turner, 2004, 56–65). This suggests that, while its incidence and impact can be significantly reduced, it is unlikely to be completely eradicated, as a result of which the zero-tolerance approach risks becoming both counterproductive and overly punitive. The appropriate policy objective, I will argue, is one of harm reduction, which involves *minimization of prevalence* with respect to the primary phenomenon combined with *attenuation of impact* for the ineliminable portion. This approach is most familiar from debates over substance abuse and drug policy, but as Weinstock and Dea have argued, it provides a perfectly general framework for thinking about normative aspects of policy in non-ideal contexts. After describing the basic principles of the approach, I will outline the specific policy implications it has for minimizing the harm caused by attitudinal racism.

[1]See, e.g., Zack (2023), Goldberg (1990), Garcia (1999, 1996).

[2]See Weir (2021). For a sophisticated defence of the "racism-is-taught" thesis, see Meltzoff and Gilliam (2024).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

1. The Philosophy of Harm Reduction

Most people who are familiar with the concept of harm reduction came to an awareness of it through debates over "supervised consumption sites" (SCS), originally developed to reduce the transmission of HIV among intravenous (IV) drug users, but since generalized to other controlled substances (Health Canada, 2024). As is well-known, possession of the drugs favored by IV users is prohibited by law and significant enforcement efforts have been undertaken to eliminate their availability. These efforts have largely failed, in the sense that there is a robust black market for these drugs, along with a relatively large group of users willing to risk arrest, and in some cases imprisonment, to purchase and consume them. Because of the illegality, users often consume these drugs in unsafe conditions. This dramatically increases the harm caused by their addiction, first by propagating bloodborne diseases (through needle-sharing), and second by increasing the risk of accidental death through overdose (by isolating users). The SCS is intended to eliminate these two harms, by providing users with clean, sterile, individual "works," and by providing medical supervision of consumption and its sequelae. The rationale is something like the following: "even though drug use is illegal, some people are still going to do it, and as long as they are doing it, we have an obligation to minimize the harm caused." This obviously contradicts the zero-tolerance approach associated with the "war on drugs," because it involves public health officials observing, failing to interdict, and in certain respects even facilitating, otherwise illegal activity.[3] At the same time, it stops short of a commitment to legalization. Many supporters of SCSs consider "hard" drugs such as heroin, methamphetamine and especially fentanyl, to be dangerous, and believe that they should remain prohibited. The harm reduction approach is targeted only at a specific class of users who are resistant to ordinary sanctions. This combination of prohibition for some and tolerance for others is not obviously coherent, and so questions have been raised about the integrity of the approach.

Both Dea and Weinstock have argued that, not only is the normative structure of the approach coherent and defensible, but it can be generalized beyond the case of IV drug use (and sex work, where there have also been initiatives along these lines) to entirely new domains (Dea, 2016; Weinstock, 2013, 2023). My own philosophical reconstruction of the policy approach is as follows. The chain of reasoning that eventuates in a call for harm reduction begins with widespread acceptance of the claim that a particular sort of behavior should not occur, based on a moral imperative (e.g., "people ought not do *x*" or "it would be best if no one did *x*,") derived from some value (Heath, 2025). The first-best response is then to prohibit *x*, in the hope that it will no longer occur, with the second-best response of punishing those who fail to comply with this prohibition. This produces what is often described as a zero-tolerance approach. Because the goal is complete prohibition, there are no circumstances in which we should be willing to observe occurrences of *x* and not intervene, interfere, or punish the perpetrator. Of course, the goal of eradication may not be achieved immediately, and so it will often be wise to persist with the punitive response, perhaps even escalating it, based on the belief that, with a bit more effort, it should be possible to prevent *x* from occurring. In time, however, and for various reasons, one might come to the conclusion that it is impossible to prevent *x* entirely. This then calls for a third-best normative response, to determine how the residual occurrences should be handled (Heath, 2025; Weinstock, 2013). The harm reduction approach constitutes one answer to the question. It focuses, as the name suggests, on reducing the harm caused by the remaining occurrences of *x*, both to the actor and third parties. These harms, it should be noted, typically figure quite prominently in the rationale for the original decision to prohibit *x*. Thus the harm reduction approach is usually motivated by the same values or moral considerations that animated the initial effort at prohibition. The difference, as Weinstock

[3]In Canada, a special ministerial exception is granted to particular sites, so that possession of controlled substances is technically not illegal in those locations. But they are illegal the moment one steps out the door onto the street, and so there is still a deliberate failure on the part of public officials to suppress illegal activity.

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

suggests, is that the zero-tolerance approach adopts a deontological stance toward these values, by imposing a categorical prohibition, while the harm reduction approach adopts a more consequentialist orientation, seeking to minimize negative effects.[4]

The crucial step in the argument involves the transition from the initial goal, of reducing prevalence to zero, to the more pragmatic goal of attenuating the harms produced by the residual portion. What makes this approach particularly controversial is not only the abandonment of efforts at prohibition for this segment of the population, but also the suspension of punishment. Both aspects stand in need of justification. With respect to the "fact of unenforceability," one can begin by cataloging the ways in which a particular form of behavior can elude prohibition. All systems of social control rely on a combination of internal motives, typically inculcated through socialization, and external ones, most obviously the threat of punishment. The power of socialization has limits, however, and some psychological dispositions (e.g., anger, jealousy, appetite, sexual arousal) prove highly resistant to elimination. There is also the simple fact that young people often rebel against heavy-handed socialization practices. Thus, the *Walden Two* scenario envisioned by B. F. Skinner (1976), in which individuals could be socialized to respect any set of constraints through appropriate conditioning, is not realizable in practice. Second, there is the fact that punishment has limits, primarily because it requires a great deal of information in order to be applied correctly, which individuals are often in a position to conceal. The option of compensating for reduced chances of apprehension by making punishments more draconian is also subject to limits, both because of a concern for proportionality and a reluctance to apply severe sanctions in cases where their deterrent effect fails.

One can see all of these limitations on social control at work in the example of drug use, since addicts experience powerful cravings that make them highly motivated to violate the prohibition; the drugs and the activity of ingestion are easily concealed (e.g., in private dwellings) making apprehension difficult; and the sanctions are not backed by very strong retributive sentiments because the behavior causes little harm to third parties. Speaking loosely, we may say that it is "impossible" to prevent certain forms of behavior from occurring, or that it is not "feasible" to prohibit it completely, but typically what we mean is simply that enforcement becomes unreasonably costly as prevalence declines, or that the punishments would need to be excessively severe, such that they would violate proportionality. With drugs, there is no doubt a large segment of the population that can be deterred through the simple threat of criminal sanction. Thus, the first 20% or 30% reduction in consumption can be achieved at practically no cost. The next 20%, however, involves by implication individuals who are not so easily deterred, and so may need evidence of active enforcement to stop. Achieving prohibition with this group will be more costly. The next 20% will be even more difficult to deter, and so on (inattention to the escalating cost of enforcement generates what Breyer, 1993 refers to as "the last 10 per cent problem" in regulation, where the attempt to achieve perfect compliance in specific cases winds up reducing overall compliance).

These considerations impose a ceiling on how large the expected disvalue of the deterrent can become. With respect to drug prohibition, there are the obvious budgetary costs of policing, and the fact that resources dedicated to the prevention of one particular class of activities are resources that are *not* dedicated to the prevention of some other. But there is also the fact that policing can be highly intrusive, and the public in many countries has limited tolerance for the violations of individual privacy and freedom that would be required in order to achieve greater suppression of the drug trade. As a result, the probability of apprehension is fairly low for many drug offences. Furthermore, unlike countries that execute people convicted of drug offences, most Western countries are quite reluctant to impose massive penalties, especially for mere possession, which

[4]"The expectation is that reasonable opponents of a controversial practice will adopt a consequentialist rather than a categorical stance with respect to their preferred value when they are made to realize that outright prohibition is unavailable," (Weinstock, 2013, 147). He describes this in terms of the distinction between *honouring* a value and *promoting* it.

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

is a primarily self-regarding action (i.e., imposing no direct harms on others).[5] The result is that a segment of the population becomes essentially undeterrable, not in principle, but relative to the enforcement system that we are actually willing to implement.

There is, it should be noted, one special case of unenforceability, which is normatively less complex than the above scenario. This involves instances in which prohibition efforts have become *counterproductive*, in the specific sense that increased effort to eradicate *x*, at the margin, is likely to increase rather than decrease the incidence of *x*. The typical case occurs when the enforcement effort itself increases the motivation, on the part of some, to engage in the prohibited behavior. This occurs most often with adolescents, who often assign positive value to acts of rebellion, risk-taking, non-conformity, or even criminality. Prohibition of an action may give it an aura of dangerousness that increases its attractiveness. Punishment of offenders may have a similar effect of turning them into role models (e.g., outlaws, martyrs). Because of this possible consequence, public health officials dealing with drug abuse must be attentive to phenomena such as "heroin chic," in order to avoid engaging in messaging that will inadvertently romanticize drug use (Mold, 2021). In such cases, it makes sense to say that the prohibition is unenforceable, because the limits of social control are at some point reached, such that further efforts at prohibition become literally self-defeating.

As Weinstock observes, one need not be a consequentialist in order to think that punishment is not the only possible response to wrongdoing; any moral framework in which the idea that an intervention might "do more harm than good" provides an intuitive basis for the construal. Most deontologists are willing to grant that there is some point at which deontic prohibitions are defeated by countervailing considerations (such as an aversion to state coercion, opposition to surveillance, a concern over proportionality in punishment, etc.). This defines what we might refer to as an "enforcement threshold," which constitutes the level at which the state becomes unwilling to further escalate punitive measures. A moral "fanatic" or "zealot" is commonly regarded as someone who is unwilling to grant that there is such a point and so continues to press for prohibition even when most people would say that the efforts have become too costly, in term of the other values that are at stake. The question then becomes whether, once the enforcement threshold has been reached, there is any point continuing to apply punishment to those for whom it is unlikely to have any deterrent effect. In some cases the answer may be yes, because we consider the punishment to be deserved, and so are willing to inflict it regardless of whether it serves any ulterior purpose. This is why we do not consider harm reduction an appropriate response to crimes such as assault and rape.[6] In other cases, however, there is a great deal less unanimity about the appropriateness of punishment or the deservingness of offenders, which gives rise to a general willingness to suspend the application of punishment, and to focus instead on attenuation of the harms caused by the behavior.

1.1. Objections to harm reduction

Politically, most of the opposition to harm reduction in drug policy comes from those who have highly retributive impulses, and so would like to see users punished "as a matter of principle" regardless of whether it serves any constructive purpose (MacCoun, 2013). Even among those who are reluctant to consider drug consumption *malum in se*, many regard the fact that it is illegal as sufficient to make users deserving of punishment. To some degree, however, this insistence on punishment is a matter of preference, since the strength of retributive sentiment is subject to considerable variation in the population. In the discursive public sphere, therefore, opponents of harm reduction typically buttress their punitive impulses with arguments intended to expose flaws in the case for harm reduction. There are three important objections to the policy approach that are commonly made.

[5]See Harm Reduction International (2023).

[6]The question of why we should not adopt a harm reduction approach to murder is one that concerns Dea (2020, 308).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

The first argument, and by far the most important, is the allegation that the two components of the harm reduction approach work at cross-purposes, insofar as the attenuation of harm is likely to *increase* prevalence through a moral hazard effect. This concern is greatest with paternalistic prohibitions such as drug control, where both the chance of arrest and the probability of adverse health impacts serve as a major deterrent to use. To the extent that an SCS eliminates both of these expected costs, it seems likely to reduce the deterrent effect of the prohibitionist policies in place and so lead to increased incidence at the margin[7] (thus, critics often describe such policies as "enabling" drug abuse). This concern, it should be noted, is exacerbated in many public debates by advocates of harm reduction who treat the prevalence of drug consumption as entirely exogenous to the calculation of costs and benefits (a view that is encouraged by the so-called disease model of addiction), or else favor legalization, and so recommend supervised consumption as an *alternative* to prohibition. Advocates for either position simply ignore the moral hazard problem, because they do not consider it a problem.

In this discussion, however, I have been adopting a philosophical formulation of the model that features a commitment to both prohibition and attenuation of harm. Theorists who endorse this approach do need to worry about moral hazard problems. The correct way to take these effects into account, however, is not to adopt a zero-tolerance attitude toward moral hazard, but rather to treat these effects as one more factor that must be taken into consideration when determining the appropriate threshold for the transition from prohibition to attenuation. If the attenuation of harm is likely to produce some increased incidence, then one should impose a somewhat higher enforcement threshold, with the recognition that there will be some backsliding over that threshold once the harm reduction policy is implemented. It is also important to recognize that the moral hazard effects are likely to be significant only with paternalistic prohibitions. Because the harms that are targeted by these policies are primarily borne by the individual who is choosing to do x, attenuation of those harms necessarily reduces the disincentive to do x and so can be expected to increase its prevalence. With other-regarding actions, by contrast, attenuating the harms may not generate any moral hazard effect at all, since the harms are borne by third parties. Because the behavior that I am concerned with in this paper primarily involves the imposition of harm on others, the concern about moral hazard is not central.

The second major objection to the harm reduction approach involves concern over the expressive effect of abandoning the commitment to complete prohibition. On the prohibition side of things, we declare a certain action to be intolerable, and yet on the attenuation side of things, we wind up tolerating it. In particular, it is difficult to formulate a harm reduction policy that does not require, in certain instances, that public officials refrain from punishing someone who is observed performing the supposedly prohibited action.[8] This can also generate a moral hazard effect by undermining the deontic force of the prohibition, reducing the stigma associated with the action, or even just by sending mixed messages about its prohibited status. But it is also subject to the more principled objection that harm reduction normalizes the behavior in a way that undermines the prohibition[9] (again, one can see this with SCSs, which are often portrayed as a Trojan horse for legalization, on the grounds that they may increase social acceptance of hard drug use).

[7]It is worth distinguishing, in this context, SCSs from "safe supply" policies (Ivsins et al., 2020). The former leaves it up to the addict to obtain drugs (illegally), it merely involves public officials ignoring this at the consumption site. Safe supply, on the other hand, involves government-supplied prescription opioids. There are reasonable grounds for concern that the latter increases prevalence. See Nguyen et al. (2024).

[8]Philosophically, one might think of this as the consequentialist orientation of the attenuation segment sitting uneasily with the deontological orientation of the prohibition and potentially undermining it.

[9]King refers to this as the "mixed-message criticism" (2020, 329).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

This argument suggests that, while it may be permissible to implement harm reduction policies *de facto*, they should never be adopted *de jure*, because of this expressive effect. The point about *de facto* adoption is important, because there are very few policy domains in which complete elimination of harmful behavior is attainable. In these areas, the limits of prohibition are usually imposed by the budgetary constraints of enforcement agencies. In each major category of crime, for instance, the police do not come anywhere near complete eradication, and so typically spread their efforts over multiple categories (which leads to the commonly voiced complaint, among those arrested for misdemeanors and public order offenses, that there are "unsolved murders" the police should be attending to). The police also, in conjunction with various social welfare agencies, offer support services to victims. What some of these amount to are harm reduction policies in everything but name. And in a sense, it also provides a response to the expressive critique, because it shows that one can adopt a harm reduction policy without necessarily publicizing that fact. But second, it shows that these policies are already common and so should not be so difficult to explain to the public in a way that will avoid mixed messaging. Furthermore, it seems unnecessarily cruel to refuse help to victims, on the grounds that this "sends the wrong message" to potential perpetrators, in cases where one has no realistic expectation of preventing those perpetrators from acting.[10]

Finally, there is the concern that adoption of a harm reduction approach will lead to premature abandonment of enforcement efforts or missed opportunities for prohibition (MacCoun, 2013, 92; Weinstock, 2013, 145). Attenuation of the harms imposed may induce complacency or sap the motivation of various constituencies that might otherwise press for elimination of the behavior. Although conceptually somewhat different from moral hazard, this is also a feedback mechanism that would have the effect of increasing prevalence of the prohibited behavior. Critics of harm reduction often express skepticism about the calculations involved in determining the enforcement threshold, based on the suspicion that more effective prohibition could be achieved through greater commitment or effort. Part of the popularity of "tough on crime" policies, for instance, is that the members of the general public tend not to see (or suffer) the cost of enforcement efforts, whereas the effects of crime are a great deal more apparent, which inclines them to believe that greater prohibition could always be achieved through tougher enforcement. Not only does attenuation of harm contribute to this perception of laxness, there is also the perception that it papers over the problem, by increasing social acceptance of the failure to achieve proper enforcement.

Again, it is important to acknowledge that this is a reasonable concern, but that it does not constitute an insuperable obstacle to the implementation of harm reduction policies. What it suggests, instead, is that the development and implementation of such policies must be grounded in high-quality social science, with real metrics and measurable data, and not just *ad hoc* or intuitive perceptions of social impacts. It also implies, however, that while public *input* will always be a crucial element in the development of such policies, decision-making with respect to their implementation generally should not be made by the broader public (for example, the American practice of having democratic elections to appoint key players in the criminal justice system—sheriffs, district attorneys, and even judges—is widely blamed for the overly punitive character of that system, because the public considers the elimination of crime to be an attainable objective and so favors a zero tolerance approach [Pfaff, 2017]). While implementation should always be subject to democratic control through high-level leadership, the specific formulation of harm reduction policies, and in particular, the specification of the recommended enforcement threshold, must necessarily be informed by expert judgment.

[10]Consider the peculiar Jane Doe case, where Toronto police failed to warn potential victims of a rapist who entered through the balcony of high-rise apartments, on the grounds that they wanted to catch him, and believed that a public warning would lead him to change his modus operandi (Doe, 2003).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

2. Racism is Ineliminable

I have spent some time offering both a precise characterization of the harm reduction approach to policy and responses to the major objections that have been raised against it, because my central objective in this paper is to recommend a more controversial application of it to the problem of attitudinal racism. I hope it will be uncontroversial to observe that the basic approach to racism in most Western societies has been to seek its complete eradication. If evidence is required, one could point to the widespread observance of "International Day for the Elimination of Racial Discrimination" (Canadian Human Rights Commission, 2023). Or one might consider the "grand challenges" initiative undertaken by the American Academy of Social Work and Social Welfare, which outlines a set of projects that its members are invited to lend their support to. Although all the goals are highly aspirational, some are nevertheless described using incrementalist language (e.g., "reduce extreme economic inequality," "promote smart decarceration") while others are formulated in absolute terms (e.g., "end homelessness," "eradicate social isolation"). The struggle against racism is formulated in absolute terms ("eliminate racism").[11]

The attainment of this objective is somewhat complicated by the fact that most governments that seek to eliminate racism are also committed to respecting a variety of liberal freedoms, including freedom of speech, and so are reluctant to employ state power to suppress the mere expression of racist beliefs and attitudes.[12] This gives rise to a division of labor between the public and private spheres, where the state has focused primarily on legal prohibition of racial discrimination (while also sometimes promoting more ambitious outcome-oriented racial equality schemes), while discouragement of racist belief, speech, and association has largely been left to the informal public sphere (which includes significant efforts undertaken in the public school system, as well as within large organizations). As a result, a great deal of the effort to control racism has involved attempts to inculcate appropriate internal motives, rather than the imposition of external coercive control, which results in a very low enforcement threshold. Intuitively, one might think that there is an internal tension between this commitment and a zero-tolerance attitude toward the primary phenomenon. The most common approach to reconciling the two seems to rest on the conviction that racism is a form of learned behavior, which should make it highly responsive to non-coercive social intervention.[13] This is usually based on a problematic inference from the observation of young children, who are often indifferent to racial distinctions. Closely tied to this is "contact theory," which in its early formulations suggested that racial biases are acquired through false representations of group characteristics and so can be eliminated merely through increased interaction among members of these groups.

This complex of ideas has become a source of considerable frustration to many social scientists who study these questions, because of its weak empirical grounding. According to Karen Stenner:

> We tend to imagine, despite a preponderance of evidence, that everyone can be socialized away from intolerance toward greater respect for difference, if only we have the will, the resources, and the opportunity to provide the right experiences. This thinking is, of course, consistent with the notion that intolerant attitudes are primarily learned; hence, they can be "unlearned." According to this wishful understanding of reality, the different can remain as different as they like, and the intolerant will eventually have their intolerance educated out of them. But all available evidence indicates that exposure to difference, talking about difference,

[11]The anti-racism goal was for some reason controversial and was not included in the original 12-point plan. "Eliminate racism," was added later, which explains the odd decision to have 13 objectives (Teasley et al., 2021). Note that the group chose *not* to endorse the goal of prison abolition, favoring instead the incrementalist goal of "promoting smart decarceration," which suggests that feasibility considerations were playing a role in their choice of language.

[12]There are important exceptions to this, such as Germany (Satariano & Schuetze, 2022).

[13]For example, "The ability to eliminate racism is predicated on the assumption that racism is learned and can be unlearned," (Buhin & Vera, 2009, 47). See also Adams et al. (2007, 60).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

> and applauding difference – the hallmarks of liberal democracy – are the surest ways to aggravate those who are innately intolerant, and to guarantee the increased expression of the predispositions in manifestly intolerant attitudes and behaviors (Stenner, 2012, 330).

Although this is a somewhat extreme view, Stenner draws two conclusion from her analysis that deserve to be taken seriously: first, that strategies for combating racism should be developed in a way that is much more responsive to evidence about what does and does not work; and second, that our goal should not be complete elimination of intolerance, but rather minimization and management for the segment of the population that remains resistant.

In a similar vein, Richard Thompson Ford has argued that the outsized role played by the judiciary in the formulation of the government response to racial segregation in the U.S. had the unfortunate effect of creating unrealistic expectations. Because the court formulated both the problem and its remedies in terms of individual rights, many concluded that reducing the incidence of racial discrimination to zero was a reasonable expectation (in the same way that, e.g., everyone's right to freedom of speech can be respected, or every criminal suspect can exercise the right against self-incrimination). As a result, the implicit benchmark against which many Americans judge the amount of progress made with respect to racism is its total elimination. If reduction of racial hostility had instead been formulated as an ordinary *policy* objective, the focus might have been on improvement over the status quo. "By this earthly standard," Ford argues, "civil rights are among the most successful policy initiatives in American history" (2005, 171). With respect to African-Americans, in particular, it is difficult to find any measure on which the climate of opinion in America has not had a consistently positive trajectory of improvement.[14] Indeed, support for certain forms of discrimination became so low in the U.S. over time that the questions pertaining to them were dropped from the General Social Survey.

Given this record, it might seem natural to expect improvement to continue until racial hostility and discrimination disappear entirely. Ultimately, of course, it is an empirical question whether they will, not something that can be answered on the basis of theory. There are, however, empirical grounds for pessimism. First, the belief that racism is sustained purely through social learning is either false or else true only in a very narrow, technical sense. While there are obviously environmental and social aspects to its specific manifestations, racism involves the activation of fairly deep-seated and universal features of human social cognition and behavior (which I shall describe below). Heuristically, one would suspect this based on its ubiquity, as well as the fact that groups who have suffered racial prejudice also have a distressing tendency to harbor their own racist attitudes. There is also the ease with which racism and intolerance can be activated, and the consistency of the tropes (e.g., the association of outsiders with disease and contagion), which suggest a persistent underlying psychological basis. And finally, there is the high level of emotional activation often associated with racist attitudes, which both makes the response difficult to suppress and suggests a deeper psychological foundation than just learned stereotypes (Garcia, 1996, 9).

I should note also that there is a tendency in contemporary discussions to expand the scope of the term "racism" to encompass many phenomena that have not traditionally been described by the concept. Many of these broader definitions make it a foregone conclusion that racism cannot be eliminated. For instance, while the meaning of terms such as "structural," "institutional," and "systemic" racism is not perfectly settled, they are often used in a way that simply picks out a racial disparity in the outcome of some social process, independent of how it was generated (Kendi, 2019, 21). It is difficult to imagine circumstances in which this sort of racism could ever be abolished. Similarly, some have tried to show that there is a necessary connection between racism and capitalism. Others have even claimed that racism is metaphysically necessary (Wilderson, 2020).

[14]See Bobo et al. (2012), 38–83. Note that the *perception* that the climate of opinion has worsened has increased in recent years (Hughes, 2024, 92).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

I will refrain from drawing upon these broader views on the grounds that they make my conclusion about ineliminability somewhat too easily arrived at. I will instead adhere to the traditional definition of racism as rooted in a psychological state and argue that even in this narrower sense of the term, there is good reason to believe that it cannot be eliminated.

There are at least three major aspects of human social cognition that are implicated in the complex of beliefs and behavioral dispositions associated with racial animosity:

2.1. Groupishness

Studies of human infants have shown that we impose a very coarse classificatory system on the world prior to the development of linguistic competence (and thus presumably with little or no cultural learning). Very young infants exhibit a complex set of responsive dispositions toward human faces, in particular, which reveal classificatory tendencies (Marquis & Sugden, 2019). There is a widely accepted line of argument, however, which suggests that it is unlikely that humans have any innate disposition to classify individuals on the basis of race, or even notable features associated with ancestry, simply because these group differences were not present in the environment of evolutionary adaptation (Cosmides et al., 2003). For over 90% of human history, people lived in small, culturally homogeneous groups that seldom exceeded 100 or so individuals. Furthermore, while most groups appear to have been exogamous, marital exchange typically occurred with neighbors who were geographically quite proximate. As a result, human infants grew up in social environments in which they were not exposed to significant racial differences, suggesting that there would have been no adaptive value to engaging in such classification.

This is the good news, as far as the eliminativist project is concerned. The bad news is that the small group environment in which humans evolved was also quite competitive, involving high levels of solidarity among insiders combined with distrust and antagonism toward outsiders. This appears to have led to the evolution of a set of cognitive and behavioral heuristics that encourage a high level of "groupishness" (Haidt, 2012; Richerson & Boyd, 2001). This involves, first, a propensity to classify individuals in the social environment into categories of "in-group" and "out-group" (or "us" and "them"); second, a disposition to act more cooperatively toward and to experience more sympathy for in-group members; and finally, a disposition to act less cooperatively, or more aggressively, toward out-group members. The cognitive and behavioral aspects of this psychological complex tend to come as a package. As Henri Tajfel and John Turner have observed, in a survey of the literature, "the mere perception of belonging to two distinct groups—that is, social categorization per se—is sufficient to trigger intergroup discrimination favoring the in-group" (2004, 56). They do not hesitate to describe this as the psychological basis of ethnocentrism, or what we might call *exophobia* (i.e., social difference aversion).

Apart from its cultural universality, the fact that such a complex suite of behaviors can be activated so reliably and easily, even in fairly young children, suggests that the basic psychological complex is not the product of cultural learning or socialization. Furthermore, once group identification has been triggered, it tends to override other potential sources of solidarity (in one fairly typical study, an explicitly random classification of individuals into groups proved to be "a more potent determinant of discrimination than perceived interpersonal similarities and dissimilarities not associated with categorization into groups" [Tajfel & Turner, 2004, 57]). These two factors combined might suggest that exophobia is an inevitable feature of the human condition, and indeed, surveying the social psychology literature on this subject can be quite discouraging. There is, however, one bright spot, which is that our evolved psychological complex appears to be agnostic about the *basis* of the classificatory scheme that triggers in-group identification. Furthermore, once a particular basis has been imposed (or becomes focal), it tends to occlude all others. As a result, our groupishness can be manipulated or spoofed (in some cases, easily so) (Voorspoels et al., 2014; also Brewer & Miller, 1984). This appears to provide the psychological grounds for an early and ubiquitous cultural adaptation, which is that all known small-scale societies adopt symbolic

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

markers, which serve to provide easily identifiable visual cues of group identity (Boyd & Richerson, 2005, 211–212). These are entirely conventional and so must be learned by children or initiates to the group.

Because of this, there is a very narrow sense in which it is correct to say that racism is learned. Since there is nothing inevitable about the use of physical traits linked to ancestry as the basis of in-group classification, American children are not born knowing that skin color, rather than hairstyle or clothing type, is the central characteristic demarcating "us" from "them" in their society. But of course once they pick up on this—and there is a wide range of different forms of social behavior that can convey this information, many of which involve no intergroup antagonism—then everything else follows quite automatically. Most importantly, no one must teach them to experience greater sympathy and concern toward in-group members, or to behave aggressively and non-cooperatively toward out-group members. In other words, no one teaches us to be exophobic, or endophilic; cultural patterns merely provide direction to these impulses.

Given this understanding of our exophobic dispositions, it is not difficult to see how certain efforts to eliminate racism can become counterproductive. The most direct strategy to reduce racial animosity is to reduce the salience of that particular classification in the social environment, so that individuals refrain from drawing the in-group/out-group distinction along racial lines (e.g., it is preferable for residents of Chicago to see themselves up as "White Sox fans" vs. "Cubs fans," rather than "white" vs. "Black"). This approach, however, runs contrary to the practice, which has recently become popular in the U.S., of encouraging heightened "race consciousness" (along with firm rejection of efforts to practice "color-blindness").[15] The suggestion is that individuals should be encouraged to continue seeing racial difference and to categorize both themselves and others on this basis, and yet should suppress the exophobic behavioral complex that is associated with it.[16] The standard argument for this position is that certain forms of racial inequality cannot be combatted without drawing attention to them, or counteracting them in a targeted way, and so a strategy that distracts from racial division will have a tendency to perpetuate inequality. If this is true, however, it is not difficult to see how it could generate a paradoxical situation, in which it is impossible to advance the struggle against racism without increased activation of the psychological dispositions that are responsible for it. Many individuals will succeed either in suppressing this or channeling it in a productive direction, but many others will not. The fact that *categorization triggers discrimination*, therefore, imposes intrinsic limits on the eliminativist program.

2.2. Authoritarianism

One of the striking features of authoritarian political movements is that the complex of attitudes they channel is so consistent across different places and times. They typically bundle together a set of superficially disparate concerns, including support for strong, decisive leadership, usually exercised by a single man; adherence to conventional morality and social norms; highly punitive attitudes toward social deviance, with a preference for forceful response; and hostility toward immigrants and outsiders.[17] This has generated the suspicion that some relatively unified, underlying personality trait may predispose individuals to exhibit this complex of attitudes.[18] Serious efforts to characterize it were undertaken in the wake of the Second World War, motivated by the effort to understand the psychological underpinnings of support for fascism (culminating in the

[15]See Aleinikoff (1991), Wingfeld (2015). A commitment to color-blindness is often condemned as a form of racism, e.g., Bonilla-Silva (2013).

[16]In some versions, only members of the majority are expected to engage in the suppression.

[17]Stenner and Haidt (2018). For example, Eric Kaufman (2016) notes that support for the death penalty was highly predictive of support for Brexit in the 2016 U.K. referendum, even though the two positions bear no obvious relation to one another.

[18]Adorno et al. (1950) was the first of many publications in this vein. See Brown (2004), Duckitt (2015).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

publication of *The Authoritarian Personality*, by Theodor Adorno, Else Frenkel-Brunswik, Daniel Levinson, and Nevitt Sanford in 1950).

It is important to emphasize at the outset that the empirical research driving these discussions is not as uncontroversial as that which informs our understanding of groupishness and exophobia. The latter phenomenon is easily elicited both in naturalistic settings and laboratory conditions, and the core findings have been extensively replicated. Personality studies, on the other hand, make extensive use of questionnaires to elicit attitudes, then use statistical methods to identify connections between these different attitudes. The correlations are usually not enormously strong, data are often open to interpretation, and there are concerns about both the validity of questionnaire responses and their connection to behavior. Early efforts, including *The Authoritarian Personality*, also suffered a number of specific methodological flaws that were widely criticized. Modern work, such as Stenner's *The Authoritarian Dynamic* (Stenner, 2012), strives to avoid these pitfalls, but it remains wedded to the basic methodological precepts of personality psychology.

With these caveats in place, one may turn to the central finding, which is that certain individuals possess a set of reactive dispositions, distinct from groupishness, that generates support for authoritarianism. More precisely, what they possess seems to be an exaggerated form of the psychological complex we all possess that forms the basis of our "norm psychology" (Sripada & Stich, 2007). This features two intrinsic motives that are activated by the observation of rule-following behavior, first a desire to conform, and second a desire to punish deviance (Sripada & Stich, 2007, 287). We can be fairly confident that these dispositions are innate, not learned, because they form part of the psychological substratum that supports the development of cultural transmission (Boyd & Richerson, 2005, 83–97). But if they are under biological control, it stands to reason that there will be variability in their strength across individuals. This suggests that some people will be "hyper-conformist" in their basic orientation: not only highly motivated to conform in their own behavior, but highly distressed by a lack of behavioral uniformity among others, and angry and punitive in their reactions. This gives them the set of reactive emotional dispositions that we associate with "intolerance," which they may or may not succeed (or even try to succeed) in controlling. For reasons that are less well understood, this sort of intolerance also appears to generate support for "strong leadership" on the conventional model (Stenner, 2012, 19).

One other important and widespread finding is that these dispositions are not always evident in everyday social interaction, but are triggered by particular social conditions that are experienced as threatening (Doty et al., 2004, 86–108). Such threats include non-conformity or lack of uniformity, either exhibited with respect to normed behavior and identity, or expressed with respect to belief and values. Lack of confidence in leadership also appears to trigger the same reaction. The major consequence of this body of research has been to suggest an exception to the basic findings of contact theory, or the view that under appropriate conditions face-to-face interaction between individuals belonging to different racial or ethnic groups will tend to reduce prejudice (Allport, 1954; Pettigrew, 1998; Pettigrew & Tropp, 2006). While there is considerable evidential support for this hypothesis, there are also important exceptions, as well as a more general concern that positive results may reflect selection bias, as people who are relatively open to difference put themselves in a position where they are more likely to experience such contact. In some cases, exposure to difference appears to generate precisely the sense of threat that triggers the authoritarian response.

Some have drawn from this the alarming conclusion that liberal-democratic societies are self-undermining, because they encourage various forms of diversity that in turn give rise to authoritarian backlash, a claim that has been met with some criticism.[19] Yet, one need not endorse this pessimistic view to see that there are nevertheless cautionary lessons to be learned. First, it suggests that certain attempts to eliminate racism, especially those involving increased contact, may become counterproductive over time. While they may initially enjoy some success, as they experience

[19]See Stenner and Haidt (2018). For critique of this view, see Dunn (2014).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

uptake from individuals who are relatively open to experience, when extended to the broader population, they may begin to provoke authoritarian backlash. This seems particularly likely with initiatives that seek to provoke "uncomfortable" conversations and encounters, based on the theory that this is necessary in order to overcome majority self-deception about racial privilege (Sue, 2015). Second, even in cases where anti-racism initiatives do not become directly counterproductive, they may still suffer from a "last 10 per cent" problem dealing with individuals who are highly susceptible to normative threat.

2.3. Stereotyping

The concept of a "racial stereotype" was introduced in order to describe false beliefs held by individuals about other racial groups, acquired through culturally transmitted representations (such as media portrayals).[20] A major virtue of intergroup contact was that it was expected to cure individuals of these stereotypes, by giving them direct exposure to members of other groups and thus an opportunity to learn the truth. Stereotyping is taken to be a major source of *bias* in interaction between members of different racial groups. Furthermore, while individuals who are overtly hostile to other racial groups will typically exhibit bias in their interactions with them and will be aware of that bias, stereotyping can lead individuals to exhibit bias without any awareness that their judgment is being influenced in this way. There has been an explosion of interest in such forms of "implicit bias," particularly with the discovery that many seem linked to subconscious associations that individuals hold (which can be elicited through an implicit association test).[21]

Over time, however, the term stereotype came to be used in a wider sense, to refer not just to false beliefs about others, but rather to any sort of generalization about groups. These generalizations are typically false when interpreted universally, since they would seldom be true of each single individual in the group, but are often true when interpreted statistically, as claims about averages or probabilities. A stereotype that is true in this sense can be referred to as *accurate* (Ryan, 2002). For example, the claim that "Asians are studious" would typically be considered a racial stereotype in America, even though it is accurate (on average, Asian students in U.S. elementary and secondary schools spend more time doing schoolwork than members of the other major census category groups) (Hsin & Xie, 2010). The harm of stereotyping occurs when people make assumptions about a specific individual based on qualities ascribed to the racial group to which that person belongs, and are either slow or unwilling to update their priors based on evidence derived from observation of the individual.

Confusingly, stereotypes are often described as a source of bias, regardless of whether they are accurate or inaccurate, even though the term bias suggests a deviation from statistically valid inference (Jussin, 2012). Many discussions of implicit bias, for example, point to the association that many people have between Black men and violent crime in the U.S., or between Indigenous youth and substance abuse in Canada. Both of these associations, however, are based on accurate statistical generalizations about these groups (Morenoff, 2005; Sikorski et al., 2019). As a result, these stereotypes have proven quite resistant to eradication, precisely because they tend to be reinforced by real-world experience. Denouncing such beliefs as racist, or as biased, as many educators and some philosophers (e.g., Basu, 2019) have been inclined to do, sets up a conflict between epistemic and moral values that is unlikely to be definitively resolved in favor of morality.

When the term "racism" is used in this expansive sense, to include epistemically valid beliefs and inferences, then it is unlikely ever to be eliminated. And yet even restricting one's attention to the epistemically invalid generalizations that people are inclined to make, which fit more closely the traditional definition of "racist," it is not clear that they can ever be suppressed entirely in the

[20]The use of the term "stereotype" in this sense is due to Lippman (1922, 89–91).

[21]Note that there has been a great deal of controversy over these claims. See Fuller (2021), also Lopez (2017a).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

population at large. For example, when dealing with a stranger, many would consider it epistemically permissible to use that person's racial group membership to establish a baseline probability (or a Bayesian prior) with respect to certain characteristics, with the expectation that this would be updated to reflect any new evidence that comes along. What we are not permitted to do is assume that the characteristic *must* be true of the individual (this would involve essentializing group differences in a way that is widely acknowledged to be racist), or to resist updating in the light of evidence. And yet these rather subtle distinctions are unlikely to be grasped by a substantial percentage of the population, much less scrupulously adhered to. As a result, it is difficult to imagine a situation in which people continue to draw the relevant group distinctions and yet refrain from overgeneralizing on their basis.

Given the centrality of frequency-detection and expectation-generation to ordinary human cognition, it is also exceedingly unlikely that the central mechanism that gives rise to stereotyping could ever be suppressed (Monteith et al., 1998). Thus, the only real way to prevent individuals from reasoning on the basis of racial stereotypes is, again, to reduce the salience of racial cues in the environment, so that individuals refrain from categorizing the regularities being observed in the environment along racial lines. As noted earlier though, this is in tension with the practice of most current anti-racism efforts, which have been sharply critical of the attempt to cultivate color-blindness in everyday social interaction. At the very least, there is a conflict between this desire to cultivate race-consciousness and the effort to deter individuals from reasoning in accordance with racial stereotypes.

2.4 The case for pessimism

The best-case scenario for the elimination of attitudinal racism would be if these habits of thought and action constituted entirely learned behavior, in which case they could be discontinued simply through a concerted effort to stop teaching them. Even here, of course, one must be cautious. There is a good case to be made that religious beliefs are also taught, and yet several attempts to eliminate them have proven unsuccessful—to the point where Enlightenment liberals have learned to accommodate themselves to their persistence. As I have tried to show in the previous three subsections, racism is closely tied to a set of cognitive heuristics that are widely thought to be innate. Of course, one must be cautious in the conclusions that one draws from this, because innate dispositions do not inevitably find expression in behavior—we can learn to suppress them, and there is no doubt that many people living in modern, pluralistic societies do just that. The problem in the case of racism is that these heuristics seem inextricably tied to other, legitimate forms of social cognition, making it difficult to police their employment even among people of good will. Most of us like the feeling of belonging to a team, club, group, or school of thought; we appreciate the sense of ease afforded by life in a close-knit community where behavioral expectations are settled and uniform; and we rely on social categorization and expectations derived from past experience to manage everyday interactions, especially with strangers. It is very difficult to draw a sharp line between harmless and harmful exercises of these dispositions.

The situation is somewhat exacerbated by the tendency, in recent years, toward adoption of increasingly expansive conceptions of racism (Blum, 2002; Carmines et al., 2011; Sniderman et al., 1991). Moving the goalposts makes it practically inevitable that racism will resist efforts at eradication. To pick just one example, the recent focus on the elimination of racial "microaggressions" makes the elimination of racism tantamount to the abolition of rudeness in interpersonal relations involving individuals of different races, which I think would strike most people as an unobtainable outcome (Sue, 2010). And yet even if one sticks to the old-fashioned conception of racism, focused on explicitly abusive speech, prejudice, and discrimination, it seems unlikely that racism will ever be completely eliminated. My argument in support of this contention is twofold.

First, there is the fact that many of the strategies that have been employed to combat racism have provoked backlash, making them directly counterproductive. For example, with respect to the

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

elimination of racial slurs, the development of a strong taboo against certain uses of language can have the perverse effect of transforming these terms into swear words. This is sometimes referred to as the "paradox of profanity," where the effort to suppress particular terms winds up promoting their use, precisely because of the censorious reaction that their use generates (Bergen, 2016, 216). One can see this at work in the effort among Christians to stop people from taking the lord's name in vain (in expressions such as "God damn it" or "*crisse de tabarnak*"). The fact that people were offended by these sacrilegious expressions is part of what made it satisfying to say them. Accordingly, as their use became increasingly common and people became less offended by them, these terms became less satisfying, as a result of which people began to shift toward words that retained the power to shock. Because of this, a more measured response, aimed at discouraging the use of certain words, rather than prohibiting them altogether, may wind up doing a better job at reducing their use in the long term.

A similar concern arises with the attempt to heighten race-consciousness, which is often seen as essential to the eradication of racial injustice, yet seems unlikely to provoke purely pro-social manifestations of groupish thinking. In particular, the attempt to persuade majority group members, such as whites in the U.S., to adopt an explicitly racial identity seems likely to become counterproductive. This is particularly important in the field of "whiteness studies," where many assert that white privilege is based, at least in part, on the fact that whiteness is an unmarked racial identity (Kolchin, 2009). And yet it is not clear that the proposed solution, which involves cultivating an explicitly racial white identity in the majority population, is going to produce greater generosity and tolerance in relations with minorities. It seems also to be provoking a resurgence of traditional white nationalism. Even if the intervention produced one white nationalist for every ten "repenters" (Bright, 2023) it clearly would not be achieving its goal.

Second, there are reasons to worry that the last 10% problem will be quite severe in the case of racism. If one imagines racial intolerance as subject to a normal distribution, the problem is that there are relatively few interventions that specifically target the tail end, where the most problematic cases lie. In part this is due to the relative success of anti-racism initiatives, which despite falling short of eliminating the problem entirely, have nevertheless succeeded in making the overt expression of racist views rather uncommon in formal settings. As a result, despite the fact that most anti-racism initiatives are not overtly punitive, those who harbor racial animus are generally loathe to self-identify, which in turn makes it difficult to focus remedial efforts on them. The majority of anti-racism initiatives seem to be focused, therefore, on further displacing the mean, presumably in the hope of reducing the number of individuals falling below a certain threshold of tolerance. Attempting to shift the entire population is obviously a very inefficient way to deal with the problems posed by outlying cases. Apart from being costly, it also runs the risk of becoming counterproductive, as individuals close to the mean are obliged to redo exercises whose point they have already fully internalized.

It is also worth noting that the commitment to freedom of expression puts rather severe constraints on the willingness of the state to employ coercive power to limit racist speech. The zero-tolerance orientation has, therefore, generated significant pressure to reduce traditional inhibitions against legal regulation of speech and has led many institutions to adopt internal norms to punish racist speech. Some U.S. universities, for example, created "bias incident" hotlines and incident reporting systems, so that episodes of racist speech could be logged and possibly subjected to disciplinary action. Much of this subsequently descended into chaos, first with the eruption of conflict over Israel/Palestine and the emergence of intractable disagreement over what constituted "hate speech" in this domain, and then with the interventions of the Trump administration. But even if the efforts had been successful, the attempt to impose a punitive regime on speech in face-to-face interaction had the effect of driving a great deal of social deviance online, where subcultures that tolerate or promote racist speech have been thriving. The overall result is that the enforcement threshold has been, and is likely to remain, relatively low for all but the most targeted and offensive

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

racist speech. Thus, the "fact of unenforceability" is likely to prevail with respect to the general prohibition on attitudinal racism.

3. Policy implications

These considerations suggest that harm reduction constitutes an appropriate policy framework for thinking about racism. Efforts should be undertaken to reduce the prevalence of these attitudes, focused primarily on the development of internal moral constraint through socialization and education, buttressed by informal social sanctions, along with legal prohibition of explicit discrimination. It should be recognized, however, that a segment of the population is likely to be resistant to these efforts and undeterred by the threatened sanctions. It should come as no surprise, therefore, that persistence in the zero-tolerance approach generates increased pressure for legal interference in domains of speech that have traditionally been protected. At some point, rather than pursuing further prohibition (e.g., by increasing the severity of sanctions, or curtailing freedom of expression), it becomes more attractive to switch to a harm reduction orientation and focus on minimizing the impact that residual hostilities are likely to have on vulnerable minorities.[22]

It is worth acknowledging at the outset an important difference between the application of a harm reduction approach to the case of opioid consumption and that of societal racism, which is that most educated people have far more permissive attitudes toward drug use, and in many cases consider consumption wrongful only because of the harm that it does to the user. With racism, on the other hand, not only is it often thought to be intrinsically wrong, but the downstream harms that it causes are imposed on innocent third parties. These two factors combine to make it seem a great deal less tolerable. And yet a central feature of the harm reduction approach is a suspension of punishment when dealing with the undeterrable segment of the population. This is likely to be the most controversial feature of my proposal, and so it is important to be very clear about what I am and am not recommending in suggesting that society move away from a zero-tolerance approach to combating racism. It is also worth keeping in mind that the harm reduction approach is still committed to reducing the overall prevalence of the wrongful attitude. It becomes distinctive only when dealing with the final, recalcitrant portion of the population, who remain unmoved by either moral suasion or social censure.

With respect to the necessity or desirability of punishment, perhaps the best model for thinking about these issues is to consider the case of racist attitudes held by minority group members themselves. Here, there is greater societal tolerance, based largely on the perception that it is less harmful, particularly when it involves disadvantaged minorities among whom own-race (or endophilic) preferences do not seem likely to generate significant opportunity-hoarding or to impose important forms of exclusion on others. (There is of course a well-known apologetic argument, which builds certain structural features into the definition of racism, resulting in the conclusion that minority group members cannot be racist.[23] Here, I am adhering to the traditional definition of racism, which focuses on the underlying psychological states. Especially given the analysis of exophobia presented in the previous section, it is not difficult to see continuity between the attitudes experienced and expressed by both minority and majority group members). At the same time, racist attitudes are still considered wrongful and efforts are often made to be discrete about the accompanying preferences (e.g., toward endogamy), keeping them confined to the private

[22]Note that unlike many supporters of the harm reduction approach, I do not consider destigmatization to be an important part of these policies, even in the case of drug abuse (see Heath, 2023, 166–168). Thus, I am not recommending that racism be destigmatized. The harm reduction approach is to be recommended primarily as a way to diminish pressure toward increased legal regulation of private life and speech.

[23]This is often summed up by the slogan that "racism = prejudice plus power" (e.g., Operario & Fiske, 1998). This view supports my position, insofar as it implies that the "prejudice" part of the equation is not what makes racism morally problematic, but rather the "power" part.

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

sphere. Nevertheless, the absence of strongly punitive reactions to racist attitudes in these cases, in which they are unlikely to produce significant harm, suggests that there is little retributive sentiment toward the attitudes taken alone. The example can therefore serve as an illustration of how it is possible to disapprove of racism (or "prejudice") without feeling that specific occurrences must always elicit a maximally punitive response.

Let us suppose then that despite the best efforts of governments, educational institutions, large organizations, and the progressive media, perhaps 10–15% of the population will remain resistant to all efforts aimed at curing them of their retrograde attitudes toward other races. Suppose further that racists have been largely purged from positions of leadership and authority in major social institutions, or failing that, the overt expression of traditional racism has been so thoroughly stigmatized that it is, for all intents and purposes, suppressed. Thus, I am imagining a scenario in which the prohibitionist stance toward racism has been largely successful, with respect to the education system, the state, the media, large corporations, etc., but has failed to eradicate the tendency completely, leaving racists an internally marginalized group. They need not constitute what Hillary Clinton notoriously referred to as a "basket of deplorables," in the sense that they are actively hateful; it is perhaps better to think of them as simply constitutionally intolerant. This underlying disposition can easily be activated in the service of various xenophobic projects, and so the primary focus of any harm reduction project, I would argue, should be the insulation of this population from the type of stimulus that is likely to produce anti-social behavior, combined with limitation of the effects of that behavior when it does arise. The tolerance of the harm reduction approach, expressed in the suspension of punishment, should be directed toward the underlying attitude, while the attenuation of impact should be focused on suppression of various forms of associated behavior.

What follows is a non-exhaustive list of specific policy proposals that might further this aim:

3.1. Tolerated self-isolation

There is a long-standing tension in liberal societies between the commitment to individual freedom, which, when combined with background pluralism, generates sorting of the population along various dimensions of difference, and the goal of promoting integration in various institutions, which is applauded on the expectation that it will produce greater tolerance and mutual understanding. Reasonably large constituencies favor policies aimed at achieving diversity (or "representation") in major social institutions, as well as with respect to work and living arrangements (Anderson, 2010). The usual objection to this essentially teleological objective is that it may interfere with individual liberty or freedom of association, prohibiting or penalizing individuals for selecting social environments that match their preferences. This has set the stage for what often seems like a contest between incompatible values or political commitments.

There is, however, an important ambiguity in the commitment to integration, which is that the goal itself provides no specification of the level at which integration must be achieved (Soon, 2023). For example, in the U.S., the concern over residential integration has typically focused on the racial composition of neighborhoods (primarily because of educational funding and access, which is local). In Singapore, by contrast, concern over racial integration extends beyond neighborhoods to encompass individual apartment buildings, each of which must respect the quotas that have been placed on the four recognized racial groups (Heath, 2023, 288–292). The latter policy obviously entails much greater restrictions on sorting, and thus more active interference with individual liberty, than the former. One way to minimize the interference is, therefore, to formulate the integration objectives at the highest level compatible with the background rationale for the policy.

There has been a trend in the U.S., however, to move in the opposite direction, imposing diversity requirements at increasingly lower levels. One can see this most clearly in popular culture, especially television and movies. Whereas it was once considered an adequate achievement for television shows or movies *in general* to feature adequate racial diversity, this has become increasingly

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

supplanted by the expectation that each television show, or every single movie, be internally diverse.[24] The concern is presumably that if racial diversity exists only within the category, racist individuals can still avoid exposure to diversity by making careful within-category selections, something that can be prevented only by introducing diversity into every single product in the category. In this case, however, the motivation for the elimination of sorting seems to be moral condemnation of the preference that gives rise to it, rather than any specific harm that it generates for others.[25] Thus, a harm reduction approach would be willing to countenance a higher degree of self-sorting, treating majority endophilia roughly in the same way that minority endophilia is treated today—permissible in the domain of private, primarily self-regarding choice, so long as it does not generate demonstrable harm to others.

3.2. Non-amplification

During the 20th century, access to public discourse was controlled by a number of very powerful gatekeepers—primarily the editors employed by major newspapers, periodicals, publishing houses, and television stations. Once a general social taboo emerged against the expression of racist attitudes in the 1970s, the result was a public sphere in which major communication channels were essentially free of overtly racist speech. This state of affairs changed dramatically with the development of the internet and social media, which had the effect of removing the gatekeepers, allowing private individuals to achieve unmediated access to potentially enormous audiences. The result was a certain measure of shock in many quarters at the level of racist speech that had traditionally been suppressed but was now finding its way into the public sphere.

The initial impulse of many has been to demand censorship of this speech, and indeed most private social media platforms have been persuaded to adopt content moderation policies that impose limits on the extremes of anti-social online behavior. But because of the commitment to free speech these policies stop short of re-establishing the older norms of decorum. Without getting into the details of these debates, one may observe that certain interventions aimed at curtailing racist speech can have the unwanted effect of amplifying it. For example, the practice of "calling out" offensive speech, as a way of expressing one's moral condemnation, also has the effect of drawing attention to it.[26] This is often accompanied by an attempt to recruit others to the effort, which can result in the original, offensive message being reproduced thousands, hundreds of thousands, or in some cases millions of times. The assumption underlying these efforts would appear to be that the punitive effect of the mass call-out will serve as a deterrent to others. And yet it is not clear that the broadcast effect will be outweighed by the deterrent effect, in such a way as to result in a net decrease in the incidence of racist speech online. The attention is itself is often rewarding to the offender, since a great deal of racist speech is rather self-evidently attention-seeking, particularly among the young. There is also a reasonable concern that the amplification may make life worse for minorities, by causing them to overestimate how common these offensive views are in the population (Hughes, 2024, 92).

One can see a similar issue with the emergence of far-right political parties, which typically cater to a combination of authoritarian and exophobic impulses. Most are aware that the expression of overtly racist views will alienate more voters than it will attract, and so maintain relatively strict message discipline in their public pronouncements. Because of this, their political opponents usually seize upon anything said that can be construed as racist and draw attention to it, in an

[24]For example, see the "Representation and Inclusion Standards" for the Best Picture award given out by the Academy of Motion Picture Arts and Sciences, https://www.oscars.org/awards/representation-and-inclusion-standards (accessed June 4, 2024).

[25]Furthermore, it seems fairly clear that imposing diversity on those who are seeking to avoid it does not produce greater tolerance per se. See Pettigrew (1998, 80).

[26]For a history of "calling out," see Ng (2022).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

effort to expose the true nature of these parties. In doing so, however, they also provide significant amplification of those attitudes. The assumption seems to be, again, that the more widely the remarks are denounced, the more effectively the attitudes will be punished, or at least silenced. Empirically, however, one might wonder whether this is so. Most obviously, it gives these parties the capacity to hijack public deliberation on practically any subject, simply by saying something inflammatory, which, in turn, shifts attention to their preferred issues.

3.3. Sublimation

Sigmund Freud believed that the various forms of instinctual renunciation required for civilization were achieved not only through straightforward repression but also through sublimation—which is to say, through the provision of substitute gratifications (Freud, 2010). This is contrary to the dominant impulse of the moralist, whose inclination is always for straightforward repression of anti-social instincts. With respect to the human taste for violence, for example, many would like to see it overcome completely, and so the prevailing compromise, in which the impulse is redirected in ways that make its expression less harmful (e.g., toward violent video games and movies), is seen as distinctly second best. Indeed, the accusation is often made that these cultural products have the opposite effect, of stimulating violent impulses that would otherwise be absent or dormant, which then risks increasing the level of genuine violence.

Again, these are complex debates that I am not in a position to adjudicate. I would like to observe, however, that the impulse toward groupishness, which appears to be central to a great deal of racial conflict, is also susceptible to sublimation through the cultivation of *other* particularistic loyalties. It is perhaps noteworthy, in this context, that the two most well-integrated institutions in American public life, namely, the military and professional sports, involve a social context that provides very strong group identity along non-racial lines. These institutions are distinctive in that they provide racially cross-cutting allegiances. Similarly, there is evidence that individuals in the U.S. who actively identify as gay, lesbian, or bisexual have an "increased propensity for interraciality in a range of interpersonal relationships" compared to heterosexuals (Horowitz & Gomez, 2018). The suggestion is that particularistic group identities crowd one another out, and so the way to combat groupishness is not to suppress it, but rather to redirect it toward less socially pernicious axes of identification (Gaertner & Dovido, 2000).

It is important also not to overlook the occasional contribution of institutions that provide higher level sources of common identity. While cosmopolitan institutions have been relatively ineffective at disrupting racial polarization, special note must be taken of the nation state and of nationalism, which possesses an affective resonance for many people that suggests satisfaction of groupish impulses. Again, just as in the case of violent entertainment, there is the suspicion that these substitute gratifications are not actually preventing the harmful expression of the attitudes being targeted, but that they may, perversely, be encouraging them. The cultivation of "tribal" loyalties, in areas such as fandom or patriotism, is sometimes seen in the same way, as offering a potentially counterproductive form of indulgence to these less-than-universalistic sentiments. A harm reduction approach, by contrast, which accepts the inevitability of such sentiments, is more likely to see these institutional strategies in a positive light, as offering a relatively harmless outlet for what are nevertheless acknowledged to be anti-social impulses.

3.4. Identity-blindness

There has been a trend in recent social justice scholarship toward increasingly vehement condemnation of so-called color-blind approaches to addressing racial inequality. During the early period of the U.S. civil rights movement, when explicit racial discrimination was rampant, it seemed relatively straightforward to assume that eliminating this sort of discrimination—so that individuals would no longer be treated differently depending on their race—would constitute an important

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

form of moral progress. At the same time, many expressed concern that the merely negative action of eliminating barriers to advancement would not be sufficient to bring about equality, and that other, more positive measures (i.e., "affirmative" action) would be required. In more recent years, however, it has been commonly asserted that the merely negative actions typically undertaken under the mantle of color-blindness were active impediments to racial progress, or worse, that they constituted a form of racist ideology. Rather than calling for supplementation of traditional non-discrimination measures, many have called for their abolition (Hughes, 2024, 26–32).

These accusations against color-blind policies are often made amidst a confusing miasma of empirical and normative definitional claims. For example, the commitment to color-blindness is often identified with the empirical belief that racial inequality has been eliminated from American society, and so nothing further need be done to rectify the problem (Neville et al., 2013, 457). Similarly, the commitment to color-blindness is sometimes equated with the view that procedures need only be facially neutral, such that disparate racial impact can be ignored entirely. These uncharitable definitions have the effect of obscuring what is, in fact, an important normative debate over the type of policies that should, or may permissibly, be adopted to combat racial discrimination.

An important recent meta-analysis of the effects of "diversity ideologies" has offered a useful reframing of the issue by drawing a distinction between "identity-blind" and "identity-conscious" approaches (Leslie et al., 2020). The authors found that, while identity-conscious "multiculturalist" approaches had the most significant effects, a combination of color-blind and meritocratic identity-blind approaches was also effective at reducing stereotyping, prejudice, and discrimination. This generates the suspicion that a great deal of the hostility toward identity-blind approaches stems from the fact that they stop short of rooting out the scourge of race prejudice; they merely prevent or prohibit individuals from acting on whatever attitudes they may have. A "meritocratic" approach, for example, which favors a rigorous emphasis on performance as a basis of advancement to the exclusion of other characteristics, reduces discrimination, but does so without requiring anyone to develop warm feelings toward others, much less affirm anyone else's identity. For this reason, it can easily be characterized as second best. And yet, if one comes to think that universal affirmation is unobtainable, this second-best outcome may begin to seem attractive.

In this context, it is important to keep in mind the extent to which institutions and policies in pluralistic societies have been improved through color-blindness. For example, the long-standing grievance of African-Americans, of not being able to hail a taxi in major cities, has disappeared from public discourse, largely due to the effects of ride-sharing apps that do not disclose the personal characteristics of riders. Similarly, there is no movement in the U.S. to restore the practice of including a photograph of job applicants on CVs, and countries like France, in which the practice remains prevalent, could probably benefit a great deal from its abolition. There are also simple steps that could be taken in residential housing and rental markets to make racial discrimination practically impossible. All of these policies would have the effect of reducing the harm produced by residual racist attitudes, by making it difficult for those who harbor such attitudes to act upon them.

3.5. Psychological resilience of victims

A common feature of minority group members who have succeeded in hostile or discriminatory environments is the psychological resilience they have exhibited in the face of adversity. In most cases, there are institutions within minority communities that bolster these efforts. In recent years, however, concerns have been raised about the possibility that certain anti-racism efforts within mainstream institutions may be inadvertently undermining the psychological resiliency that minorities require in order to succeed under adversity (Campbell & Manning, 2018; Lukianoff & Haidt, 2018). Two factors have been identified that may be contributing to this effect. First, in social environments where content-based restrictions on free speech are prohibited, it has become common to argue that racially abusive speech should be prohibited on the grounds that it is directly

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

harmful, or that it threatens the "safety" of minority group members. Although the initial motivation for such claims was presumably to call for restrictions on speech while formally maintaining adherence to traditional free speech norms, the practice of claiming that speech produces psychological harm appears to have become self-fulfilling in certain cases. Second, the zero-tolerance approach to racism has encouraged individuals to call for third-party intervention in response to even very minor incidents (Miller et al., 2018). The result has been a certain tendency toward catastrophization, as individuals are encouraged to claim psychological injury and seek redress in response to minor slights. The result, as Greg Lukianoff and Jonathan Haidt (2018) allege, is a type of "reverse cognitive-behavioural therapy," wherein our moral commitments encourage individuals to act in a way that is diametrically opposed to the habits that encourage psychological resilience.

Again, it is important to be clear about the scope of this claim. I am not recommending that minorities develop psychological resilience in the face of actual discrimination or threats. The issue concerns primarily the response to the *expression* of negative attitudes, up to and including the use of racial slurs. Advocating psychological resilience in the face of such behavior is fraught, because of the moral judgment that the incidents that require it are contranormative and so should not be happening at all. But once it is accepted that they will never be eliminated entirely, it becomes important to minimize the harm that they do. One way of reducing that harm is to minimize their psychological effects on victims, for example, to ensure that they do not cause significant psychological damage or trauma. This seems particularly achievable with respect to the category of offenses currently described as microaggressions, many of which constitute facially neutral remarks that are targeted for prohibition because of the alleged psychological harm that they cause (Rini, 2020; Sue, 2010). In such cases, it seems clear that the imperatives of harm reduction may impose burdens of adjustment on both speaker and listener.

4. Conclusion

It is not difficult to anticipate the objections that will be made to the approach to attitudinal racism being advanced here: that it evinces an overly tolerant attitude toward racist beliefs; that it is defeatist, prematurely accepting the inevitability of a state of affairs that can still be ameliorated; that it "sends the wrong message" to racists, by suggesting that their views are less damaging than they in fact are; and that by reducing the cost to society of racial animosity it may even promote greater prevalence of those attitudes. The ease with which these objections can be anticipated is due to the fact that they are not specific to the case, but constitute rather *general objections to the harm reduction approach.* They are, in other words, merely different instantiations of the same general concerns that have been raised by opponents of supervised consumption sites for drug users. That some people will raise concerns in both cases is unexceptionable—the harm reduction approach is, after all, controversial, and involves a difficult balancing act between deontological and consequentialist impulses. What I would take exception to, however, are those who would advance these objections against the harm reduction approach to societal racism and yet believe that they can be successfully addressed in the case of drug abuse. Those who maintain this stance owe some account of how the two cases differ, such that the objections have force with respect to one issue that they lack with the other.

Among those who are sympathetic to the harm reduction approach, but may have doubts about the wisdom of applying it to the case of racism, an important question will be how close they believe the causal connection is between racist attitudes and the various negative outcomes that we associate with societal racism.[27] The question, in other words, is whether stopping short of complete

[27]As Dea observes, the harm reduction approach is only appropriate with respect to "behaviours that commonly produce a range of harms that can be reduced even though the primary behaviours persist" (2020, 307).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

prohibition—allowing some people to harbor intolerant views in private—will impede the attainment of socially valued objectives, such as racial equality in the workplace or in various domains of cultural expression. The tightness of this connection will, in effect, determine whether it is possible to significantly reduce the *harms* of racism without eliminating its underlying causes. For example, how much residual racial animosity is compatible with relative equality in various dimensions of achievement, such as educational attainment or labor-market returns? My own view is that there is a great deal more leeway here than many commentators imagine. For instance, callback studies have shown that the level of labor-market discrimination suffered by Blacks in many Western countries (including Canada) is comparable to that experienced by Asians.[28] Indeed, many studies show surprisingly high levels of racial animosity directed toward Asians—surprising precisely because it has not prevented these groups from achieving high levels of income, educational attainment, residential integration, and so on.

My inclination is to think that pessimism about the possibility of achieving racial equality without eradicating racism is mistaken, precisely because our folk-sociological inclination is to overestimate the isomorphism between individual attitudes and social outcomes. In other words, there is a persistent tendency to assume too much intentionality in the processes that result in macro-level phenomena. Thomas Schelling's "segregation model" (1978) of residential housing dynamics was intended to show precisely how difficult it is to draw conclusions about individual preferences from the observation of large-scale patterns. There is considerable free play between the distribution of various traits at the individual level and the patterns that emerge at the social level. This suggests, in turn, that it may be possible to achieve racial equality in all of the most meaningful domains of life without curing everyone of their private animosities, but merely through intelligent institutional design aimed at preventing these animosities from impacting the lives and well-being of those toward whom they are directed

Acknowledgments. I would like to thank Daniel Weinstock, Andrew Potter, Brendan De Kenessey, Shannon Dea, and Abraham Singer for helpful comments, along with participants in a workshop held at the Department of Philosophy at Vanderbilt University. This paper was given as the annual Canadian Journal of Philosophy Distinguished Lecture at the annual conference of the Canadian Philosophical Association in 2024.

Joseph Heath is a Professor in the Department of Philosophy at the University of Toronto.

References

Adams, M., Bell, L. A. & Griffin, P. (Eds.) (2007). *Teaching for social justice and diversity* (2nd ed.). Routledge.

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D., & Sanford, N. (1950). *The authoritarian personality*. Harper & Brothers.

Aleinikoff, T. A. (1991). A case for race-consciousness. *Columbia Law Review*, 91(5), 1060–1125.

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

Anderson, E. (2010). *The imperative of integration*. Princeton University Press.

Basu, R. (2019). The wrongs of racist beliefs. *Philosophical Studies*, 176(9), 2497–2515.

Bergen, B. K. (2016). *What the F*. Basic Books.

Blum, L. (2002). Racism: What it is and what it isn't. *Studies in Philosophy and Education*, 21(3), 203–218.

Bobo, L. D., Charles, C. Z., Krysan, M., & Simmons, A. D. (2012). The real record on racial attitudes. In P. V. Marsden. *Social trends in American life: Findings from the general social survey since 1972* (pp. 38–83). Princeton University Press.

Bonilla-Silva, E. (2013). *Racism without racists*. Rowman & Littlefield.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford University Press.

Boyd, R., & Richerson, P. (2005). *Not by genes alone*. University of Chicago Press.

Brewer, M. B., & Miller, N. (1984). Beyond the contact hypothesis: Theoretical perspectives on desegregation. In N. Miller & M. B. Brewer (Eds.), *Groups in contact* (pp. 281–302). Academic Press.

Breyer, S. (1993). *Breaking the vicious circle*. Harvard University Press.

Bright, L. K. (2023). White psychodrama. *Journal of Political Philosophy*, 31(2), 198–221.

[28]See Quillian et al. (2019, 480). Nevertheless, commentators often assume that lower Black income levels must be caused by discrimination of the type uncovered in such studies (e.g., Lopez, 2017b).

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

Brown, R. (2004). The authoritarian personality and the Organization of Attitudes. In J. T. Jost & J. Sidanius (Eds.), *Political psychology* (pp. 45–85). Psychology Press.
Buhin, L., & Vera, E. M. (2009). Preventing racism and promoting social justice: Person-centered and environment-centered interventions. *Journal of Primary Prevention*, 30(1), 43–59.
Campbell, B., & Manning, J. (2018). *The rise of victimhood culture*. Palgrave Macmillan.
Canadian Human Rights Commission. (2023) Eliminating racism is our collective responsibility. https://www.chrc-ccdp.gc.ca/resources/newsroom/eliminating-racism-our-collective-responsibility.
Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 40, pp. 193–236). Elsevier.
Carmines, E. G., Sniderman, P. M., & Easter, B. C. (2011). On the meaning, measurement, and implications of racial resentment. *The Annals of the American Academy of Political and Social Science*, 634(1), 98–116.
Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Science*, 7(4), 173–179.
Dea, S. (2016). A harm reduction approach to abortion. In S. Stettner (Ed.). *Without apology* (pp. 317–32). Athabasca University Press.
Dea, S. (2020). Toward a philosophy of harm reduction. *Health Care Analysis*, 28(4), 302–313.
Doe, J. (2003). *The story of Jane Doe: A book about rape*. Random House.
Doty, R. M., Peterson, B. E., & Winter, D. G. (2004). Threat and authoritarianism in the United States, 1978–1987. In J. T. Jost & J. Sidanius (Eds.), *Political psychology* (pp. 86–108). Psychology Press.
Duckitt, J. (2015). Authoritarian personality. In J. D. Wright (Ed.), *International Encyclopedia of the social and Behavioral sciences* (2nd ed., pp. 255–261). Elsevier.
Dunn, K. (2014). Authoritarianism and intolerance under autocratic and democratic regimes. *Journal of Social and Political Psychology*, 2(1), 220–241.
Ford, R. T. (2005). *Racial culture: A critique*. Princeton University Press.
Freud, S. (2010). *Civilization and its discontents*, Trans. James Strachey. W. W. Norton.
Fuller, T. (2021). Do implicit racial biases have significant discriminatory effects? *Ergo*, 8, 22. https://doi.org/10.3998/ergo.1159.
Gaertner, S. L., & Dovido, J. F. (2000). *Reducing intergroup bias*. Psychology Press.
Garcia, J. L. A. (1996). The heart of racism. *Journal of Social Philosophy*, 27(1), 5–45.
Garcia, J. L. A. (1999). Philosophical analysis and the moral concept of racism. *Philosophy and Social Criticism*, 25(5), 1–32.
Goldberg, D. T. (Ed.) (1990). *Anatomy of racism*. University of Minnesota Press.
Haidt, J. (2012). *The righteous mind*. Random House.
Harm Reduction International. (2023). *The death penalty for drug offences: Global overview 2022*. Harm Reduction International.
Health Canada. (2024). Supervised consumption sites explained: Types of sites and services. https://www.canada.ca/en/health-canada/services/substance-use/supervised-consumption-sites/explained.html.
Heath, J. (2014). *Enlightenment 2.0*. HarperCollins.
Heath, J. (2023). *Cooperation and social justice*. University of Toronto Press.
Heath, J. (2025). The enforcement threshold approach to harm reduction [unpublished manuscript].
Horowitz, A. L., & Gomez, C. J. (2018). Identity override: How sexual orientation reduces the rigidity of racial boundaries. *Sociological Science*, 5(28), 669–693.
Hsin, A., & Xie, Y. (2010). Explaining Asian Americans' academic achievement advantage over whites. *Proceedings of the National Academy of Sciences USA*, 111(23), 8416–8421.
Hughes, C. (2024). *The end of race politics*. Penguin.
Ivsins, A., Boyd, J., Beletsky, L., & McNeil, R. (2020). Tackling the overdose crisis: The role of safe supply. *International Journal of Drug Policy*, 80, 102769.
Jussin, L. (2012). *Social perception and social reality*. Oxford University Press.
Kaufman, E. (2016). It's NOT the economy, stupid: Brexit as a story of personal values. *British Politics and Policy at LSE*, July 7. https://blogs.lse.ac.uk/politicsandpolicy/personal-values-brexit-vote/.
Kendi, I. X. (2019). *How to be an anti-racist*. Penguin.
King, N. B. (2020). Harm reduction: A misnomer. *Health Care Analysis*, 28(4), 324–334.
Kolchin, P. (2009). Whiteness studies. *Journal de la société des américanistes*, 95(1), 117–163.
Leslie, L. M., Bono, J. E., Kim, Y., & Beaver, G. R. (2020). On melting pots and salad bowls: A meta-analysis of the effects of identity-blind and identity-conscious diversity ideologies. *Journal of Applied Psychology*, 105(5), 453–471.
Lippman, W. (1922). *Public opinion*. Harcourt, Brace and Co.
Lopez, G. (2017a). For years, this popular test measured anyone's racial bias. But it might not work after all. *Vox*, March 7. https://www.vox.com/identities/2017/3/7/14637626/implicit-association-test-racism.
Lopez, G. (2017b). Study: Anti-black hiring discrimination is as prevalent today as it was in 1989. *Vox*, September 18.
Lukianoff, G., & Haidt, J. (2018). *The coddling of the American mind*. Penguin.

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

MacCoun, R. J. (2013). Moral outrage and opposition to harm reduction. *Criminal Law and Philosophy*, 7(1), 83–98.
Marquis, A. R., & Sugden, N. A. (2019). Meta-analytic review of infants' preferential attention to familiar and unfamiliar face types based on gender and race. *Developmental Review*, 53(3), 100868.
Meltzoff, A. N., & Gilliam, W. S. (2024). Young children and implicit racial biases. *Daedalus*, 153(1), 65–83.
Miller, R. A., Guida, T., Smith, S., Ferguson, S. K., & Medina, E. (2018). Free speech tensions: Responding to bias on college and university campuses. *Journal of Student Affairs Research and Practice*, 55(1), 27–39.
Mold, A. (2021). Just say know: Drug education and its publics in 1980s Britain. *International Journal of Drug Policy*, 88(6357), 103029.
Monteith, M. J., Sherman, J. W., & Devine, P. G. (1998). Suppression as a stereotype control strategy. *Personality and Social Psychology Review*, 2(1), 63–82.
Morenoff, J. D. (2005). Racial and ethnic disparities in crime and delinquency in the United States. In M. Ruter & M. Tienda (Eds.), *Ethnicity and causal mechanisms* (pp. 139–173). Cambridge University Press.
Neville, H. A., Awad, G. H., Brooks, J. E., Flores, M. P. & Bluemel, J. (2013). Color-blind racial ideology: Theory, training, and measurement implications in psychology. *American Psychologist*, 68(6), 455–466.
Ng, E. (2022). *Cancel culture: A critical analysis*. Palgrave Macmillan.
Nguyen, H. V., Mital, S., Bugden, S. & McGinty, E. E. (2024). British Columbia's safer Opiod supply policy and opioid outcomes, *JAMA Internal Medicine*, 184(3), 256–264.
Operario, D., & Fiske, S. T. (1998). Racism equals power plus prejudice: A social psychological equation for racial oppression. In J. L. Eberhardt & S. T. Fiske (Eds.), *Confronting racism* (pp. 33–53). Sage.
Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49(1), 65–85.
Pettigrew, T. F., & Tropp, L. A. (2006). A meta-analytical test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783.
Pfaff, J. F. (2017). *Locked in*. Basic Books.
Quillian, L., Heath, A., Pager, D., Midtboen, A. H., Fleischmann, F., & Hexel, O. (2019). Do some countries discriminate more than others? Evidence from 97 field experiments of racial discrimination in hiring. *Sociological Science*, 6(2041), 467–496.
Richerson, P. J., & Boyd, R. (2001). The evolution of subjective commitment to groups: A tribal instincts hypothesis. In R. M. Nesse (Ed.), *Evolution and the capacity for commitment* (pp. 186–220). Russell Sage Foundation.
Rini, R. (2020). *The ethics of microaggressions*. Routledge.
Ryan, C. S. (2002). Stereotype accuracy. *European Review of Social Psychology*, 13(1), 75–109.
Satariano, A., & Schuetze, C. F. (2022). Where online hate speech can bring the police to your door. *New York Times*, September 23.
Schelling, T. (1978). *Micromotives and macrobehavior*. W. W. Norton.
Sikorski, C., Leatherdate, S., & Cooke, M. (2019). Tobacco, alcohol and marijuana use among indigenous youth attending off-reserve schools in Canada: Cross-sectional results from the Canadian student tobacco, alcohol and drugs survey. *Health Promotion and Chronic Disease Prevention in Canada*, 39(6/7), 207–215.
Skinner, B. F. (1976). *Walden two*. Hackett.
Sniderman, P. M., Piazza, T., Tetlock, P. E., & Kendrick, A. (1991). The new racism. *American Journal of Political Science*, 35(2), 423–447.
Soon, V. (2023). Sorting and the ecology of freedom of association. *Journal of Political Philosophy*, 31(4), 411–432.
Sripada, C. S., & Stich, S. (2007). A framework for the psychology of norms. In P. Carruthers (Ed.), *The innate mind* (Vol. 2, pp. 280–301). Oxford University Press.
Stanovich, K. (2004). *The robot's rebellion*. University of Chicago Press.
Stenner, K. (2012). *The authoritarian dynamic*. Cambridge University Press.
Stenner, K., & Haidt, J. (2018). Authoritarianism is not a momentary madness, but an eternal dynamic within Liberal democracies. In C. Sunstein (Ed.), *Can it happen here?* (pp. 175–220). HarperCollins.
Sue, D. W. (2010). *Microaggressions in everyday life*. John Wiley.
Sue, D. W. (2015). *Race talk and the conspiracy of silence*. Wiley.
Tajfel, H., & Turner, J. (2004). An integrative theory of intergroup conflict. In M. J. Hatch & M. Schultz (Eds.), *Organizational identity: A reader* (pp. 56–65). Oxford University Press.
Teasley, M. L., McCarter, S., Woo, B., Conner, L. R., & Spencer, M. S. (2021). Grand challenge: Eliminate racism. *Grand Challenges for Social Work Initiative, Working Paper*, 26.
Voorspoels, W., Bartlema, A., & Vanpaemel, W. (2014). Can race be erased? A pre-registered replication study. *Frontiers of Psychology*, 5, 1035.
Weinstock, D. (2013). So, are you still a philosopher? In *Trudeau foundation papers, Vol.* 5. Fondation Trudeau.
Weinstock, D. (2020). Disagreement, unenforceability, and harm reduction. *Health Care Analysis*, 28(4), 314–323.
Weinstock, D. (2023). Multiculturalism as harm reduction. *Res Publica*, 29(4), 611–627.

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press

Weir, K. (2021). Raising anti-racist children. *Monitor on Psychology*, 52(4), 52–61.
Wilderson, F. B. (2020). *Afropessimism*. W. W. Norton.
Wingfeld, A. H. (2015). Color blindness is counterproductive. *The Atlantic*, September 13.
Zack, N. (2023). *Philosophy of race* (2nd ed.). Palgrave.

Cite this article: Heath, J. 2024. A Harm Reduction Approach to Attitudinal Racism. *Canadian Journal of Philosophy* 54: 482–506, doi:10.1017/can.2025.10017

https://doi.org/10.1017/can.2025.10017 Published online by Cambridge University Press