Words of Hate, Streets of Rage: Reforming the England and Wales Stirring up Hatred Offenses

Alexander Brown*

Britain has a long history of individuals using words to provoke enmity between different sections of society and of ensuing riots and other breaches of the peace. Since 1965, Britain has also had statutory offenses relating to the stirring up of hatred on grounds of race and some other protected characteristics, set forth as part of the legislative response to the aforementioned problems. This Article tests the England and Wales stirring up hatred offenses against two First Amendment doctrines. In particular, it investigates the prospects for mitigating the viewpoint discrimination manifest in these laws and whether such mitigations could render the laws overbroad. It also examines the problems of courts applying an ordinary meaning approach to the phrase "stir up." In addition, it seeks to make sense of the current sentencing guidelines for the stirring up hatred offenses having to do with the aggravating factors of speakers holding "positions of trust, authority, or influence" and audiences being "vulnerable/impressionable."

Plotting a pathway through these issues, the Article makes an appeal not simply to public order considerations but also to the value of substantive rather than formal autonomy. The goal of protecting substantive autonomy implies that one person's choice of words or behavior should not diminish the autonomy of another person, such as by circumventing or curtailing their independent, rational deliberations. In a giant leap forward for law in this area, this Article recommends a series of significant reforms to the England and Wales stirring up hatred offenses.

I. INTRODUCTION

Not for the first time in modern British history, violent protests, mass brawls, riots, and assaults involving warring sections of society cast a pall over the summer of 2024. The England and Wales stirring up hatred offenses are part of a family of public order offenses that law enforcement

* © 2025 Alexander Brown, Associate Professor of Political and Legal Theory, School of Politics, Philosophy, and Area Studies, University of East Anglia (UEA), U.K.

authorities can, and often do, call upon in dealing with people who stoke up enmities in a manner likely to contribute to such events. Yet, the England and Wales stirring up hatred offenses have also been criticized by legal scholars (as well as by Elon Musk) on various fronts.[1] Notable areas of criticism include: the reliance on the ambiguous statutory phrase "threatening, abusive, or insulting words or behavior"; the drawbacks of adopting an either/or approach to intention and likelihood elements; the need for more robust protections of freedom of expression built into the legislation as written; the narrow list of protected characteristics; the arbitrariness of how different protected characteristics are handled with different statutory phrases, elements, and free speech protections; the

1. *See* David R. Fryer, *Group Defamation in England*, 13 CLEV.-MARSHALL L. REV. 33 (1964); D.G.T. Williams, *Racial Incitement and Public Order*, 320 CRIM. L. REV. (1966); Anthony F. Dickey, *English Law and Incitement to Racial Hatred*, 9 RACE 311 (1968); Anthony F. Dickey, *Prosecutions Under the Race Relations Act 1965 s. 6 (Incitement to Racial Hatred)*, 489 CRIM. L. REV. (1968); Richard P. Longaker, *The Race Relations Act of 1965: An Evaluation of the Incitement Provision*, 11 RACE 125 (1969); P. M. Leopold, *Incitement to Hatred: The History of a Controversial Criminal Offence*, PUB. L. 389 (1977); Roger Cotterell, *Prosecuting Incitement to Racial Hatred*, PUB. L. 378 (1982); GEOFFREY BINDMAN, *Incitement to Racial Hatred*, NEW L.J. 229 (1982); JONATHAN GEWIRTZ, *The Case for Group Libel Law in Great Britain*, in MINORITIES: COMMUNITY AND IDENTITY (C. Fried ed., 1983); Dexter Dias, *A Licence to Hate: Incitement to Racial Hatred and the Public Order Act of 1986*, SOC. LAWYER 20 (1987); Geoffrey Bindman, *What Happened to Racial Incitement*, 87 L. SOC. GAZETTE 25 (1990); Geoffrey Bindman, *Outlawing Hate Speech*, 89 LAW SOCIETY'S GAZETTE 17 (1992) [hereafter *Outlawing Hate Speech*]; TARIQ MODOOD, *Muslims, Incitement to Hatred and the Law, in Liberalism, Multiculturalism, and Toleration*, in LIBERALISM MULTICULTURALISM AND TOLERATION (J. Horton ed., 1993); Anne Twomey, *Laws Against Incitement to Racial Hatred in the United Kingdom*, 1 AUSTRL. J. HUM. RTS. 235 (1994); Paul Kearns, *The Occlusion of Opinion: Incitement to Religious Hatred*, 59 AMICUS CURIAE 20 (2005); Ivan Hare, *Crosses, Crescents and Sacred Cows: Criminalising Incitement to Religious Hatred*, PUB. L. 521 (2006); Kay Goodall, *Incitement to Religious Hatred: All Talk and No Substance?*, 70 MODERN. L. REV. 89 (2007); David Nash & Chara Bakalis, *Incitement to Religious Hatred and the "Symbolic": How Will the Racial and Religious Hatred Act 2006 Work?*, 28 LIV. L. REV. 349 (2007); Alexander Brown, *The Racial and Religious Hatred Act 2006: A Millian Response*, 11 CRITICAL REV. INT'L. SOC. & POL. PHIL. 1 (2008) [hereafter T*he Racial and Religious Hatred Act 2006*]; KAY GOODALL, *Challenging Hate Speech: Incitement to Hatred on Grounds of Sexual Orientation in England, Wales and Northern Ireland,* in PROTECTION OF SEXUAL MINORITIES SINCE STONEWALL: PROGRESS AND STALEMATE IN DEVELOPED AND DEVELOPING COUNTRIES (F.C.W. Chan ed., 2010).; Alexander Brown, *The "Who?" Question in the Hate Speech Debate: Part 1: Consistency, Practical, and Formal Approaches*, 29 CAN. J. L. & JUR. 275 (2016) [hereafter *The "Who?" Question in the Hate Speech Debate: Part 1*]; Alexander Brown, *The "Who?" Question in the Hate Speech Debate: Part 2: Functional and Democratic Approaches*, 30 CAN. J. L. & JUR. 23 (2017) [hereafter *The "Who?" Question in the Hate Speech Debate: Part 2*]; JEN NELLER, *Hate Speech Law and Equality: A Cautionary Tale for Advocates of "Stirring up Gender Hatred" Offences*, in TOWARDS GENDER EQUALITY IN LAW (G. Guney et al. eds., 2022). *See also* Elon Musk's comments on X concerning the application of the stirring up hatred offenses to comments posted on social media: "The Woke Stasi", @Elonmusk, X (Aug. 8, 2024).

question of whether the offenses were necessary given that, when they were created by the Race Relations Act 1965, law enforcement authorities could already seek to prosecute a person for bringing a class of Her Majesty's subjects into contempt in the eyes of the public under the common law offense of seditious libel, and/or prosecute a person for using threatening, abusive, or insulting words or behavior with intent or likelihood of provoking a breach of the peace under Section 5 of the Public Order Act 1936; the related question of whether it is in fact significantly easier for prosecutors to prove intent or likelihood of stirring up hatred than to prove intent or likelihood of provoking a breach of the peace; and the dangers of censorious and repressive applications of these laws to areas of public discourse where free speech ought to be unmolested.

This Article discusses in detail several additional, hitherto underexplored issues related to the England and Wales stirring up hatred offenses: First, the extent to which the offenses involve viewpoint discrimination; second, the risk that avoiding or mitigating viewpoint discrimination could render the offenses overbroad (for example, reducing viewpoint discrimination could mean expanding the scope of the offenses to cover additional protected characteristics and/or the stirring up of extra emotions, sentiments, or attitudes); third, whether there are any unique problems with courts taking an ordinary meaning approach to the phrase "stir up"; and fourth, understanding and justifying the current sentencing guidelines for these offenses, which include, amongst other things, aggravating factors having to do with the speaker holding a "position of trust, authority, or influence" and the audience being "vulnerable/impressionable." Furthermore, I argue that there is a golden thread connecting each of these different issues, namely the value of substantive autonomy, as in, protecting the autonomy of those on the receiving end of words or behavior intended to stir up of extreme negative emotions, sentiments, or attitudes against other persons or groups of people on grounds of protected characteristics.

Working through these issues and drawing on the value of substantive autonomy, along with some other important rationales, I recommend several reforms to the England and Wales stirring up hatred offenses, in the shape of my proposal for a model stirring up offense, as well as reforms to the current sentencing guidelines. Taken together, these reforms are as follows: that the list of protected characteristics be expanded; that for all protected characteristics the offenses should be standardized, in particular to cover *only* the use of threatening words or behavior, and to include a conjunction of the elements of "intent to stir up

hatred" *and* "likely to stir up hatred"; that the offenses be widened to cover not only the stirring up of hatred but also the stirring up of any extreme negative emotions, sentiments, or attitudes against other persons or groups of people based on protected characteristics; that the offenses be enlarged to also cover incitement to discrimination or violence; that the offenses be given extraterritorial application; and that the sentencing guidelines be amended to provide concrete illustrations of a speaker holding a "position of trust, authority, or influence" and of an audience being "vulnerable/impressionable," specifically, the examples of political figures and the congregations, followers, or students of religious leaders or teachers.

Initially, however, I need to provide some further clarifications and background details. First, I refer at times to the conduct of stirring up hatred as being proscribable and ask what justifies it being so. By "proscribable" I mean that there is a *pro tanto* reason to proscribe. I do not mean that, all things considered, it should be proscribed. Manifold considerations (legal, moral, social, political) go into the latter sort of overall assessment.[2] In fact, there is a vast literature setting out numerous arguments both for and against various kinds of hate speech laws.[3] This

---

2. ALEXANDER BROWN, HATE SPEECH LAW: A PHILOSOPHICAL EXAMINATION 3 (2015) [hereafter HATE SPEECH LAW].

3. *See* Mari Matsuda, *Public Response to Racist Speech: Considering the Victim's Story*, 87 MICH. L. REV. 2320 (1989); Rodney Smolla, *Academic Freedom, Hate Speech, and the Idea of a University*, 53 L. & CONTEMP. PROBS. 195 (1990); Nadine Strossen, *Regulating Racist Speech on Campus: A Modest Proposal*, 1990 DUKE L.J. 484 (1990); Richard Delgado, *Campus Antiracism Rules: Constitutional Narratives in Collision*, 85 NW. U. L. REV. 343 (1991); Suzanna Sherry, *Speaking of Virtue: A Republican Approach to University Regulation of Hate Speech*, 75 MINN. L. REV. 933 (1991); SANDRA COLIVER, *Hate Speech Laws: Do They Work?*, in STRIKING A BALANCE: HATE SPEECH, FREEDOM OF EXPRESSION AND NON-DISCRIMINATION (S. Coliver ed., 1992); Charles Lawrence III, *Cross Burning and the Sound of Silence: Anti-Subordination Theory and the First Amendment*, 37 VILL. L. REV. 787 (1992); MARI J. MATSUDA, CHARLES R. LAWRENCE III, RICHARD DELGADO, & KIMBERLÉ W. CRENSHAW, WORDS THAT WOUND: CRITICAL RACE THEORY, ASSAULTIVE SPEECH, AND THE FIRST AMENDMENT (M. Matsuda et al. eds., 1993); CASS SUNSTEIN, DEMOCRACY AND THE PROBLEM OF FREE SPEECH (1993); Henry Louis Gates Jr., Anthony P. Griffin, Donald E. Lively, & Nadine Strossen, SPEAKING OF RACE, SPEAKING OF SEX: HATE SPEECH, CIVIL RIGHTS AND CIVIL LIBERTIES (1995); LAURA LEDERER, THE PRICE WE PAY: THE CASE AGAINST RACIST SPEECH, HATE PROPAGANDA, AND PORNOGRAPHY (L. Lederer & R. Delgado eds., 1995); Larry Alexander, *Banning Hate: Speech and the Sticks and Stones Defense*, 13 CONST. COMMENT. 71 (1996); Richard Delgado and Jean Stefancic, *Ten Arguments Against Hate Speech Regulation: How Valid?*, 23 N. KY. L. REV. 475 (1996); OWEN FISS, THE IRONY OF FREE SPEECH: LIBERALISM DIVIDED (1996); STEVEN SHIFFRIN, DISSENT, INJUSTICE, AND THE MEANINGS OF AMERICA (1999); Nadine Strossen, *Incitement to Hatred: Should There Be a Limit?*, 25 S. ILL. UNIV. L.J. 243 (2001); JAMES WEINSTEIN, *Hate Speech, Viewpoint Neutrality, and the American Concept of Democracy, in* THE BOUNDARIES OF FREEDOM OF EXPRESSION AND ORDER IN A DEMOCRATIC SOCIETY (T. Hensley ed., 2001) [hereafter *Hate Speech, Viewpoint Neutrality,*

and the American Concept of Democracy]; ALEXANDER TSESIS, DESTRUCTIVE MESSAGES: HOW HATE SPEECH PAVES THE WAY FOR HARMFUL SOCIAL MOVEMENTS (2002); J. Angelo Corlett & Robert Francescotti, *Foundations of a Theory of Hate Speech*, 48 WAYNE L. REV. 1071 (2002); Richard Delgado & Jean Stefancic, UNDERSTANDING WORDS THAT WOUND (2004); JON B. GOULD, SPEAK NO EVIL: THE TRIUMPH OF HATE SPEECH REGULATION (2005); Bhikhu Parekh, *Hate Speech: Is There a Case for Banning?*, 12 PUB. POL. RES. 213 (2005); Eric Heinze, *Viewpoint Absolutism and Hate Speech*, 69 MODERN L. REV. 543 (2006) [hereafter *Viewpoint Absolutism and Hate Speech*]; Steven J. Heyman, FREE SPEECH AND HUMAN DIGNITY (2008); C. EDWIN BAKER, *Autonomy and Hate Speech*, *in* EXTREME SPEECH AND DEMOCRACY (I. Hare & J. Weinstein eds., 2009) [hereafter *Autonomy and Hate Speech*]; Raphael Cohen-Almagor, *Holocaust Denial Is a Form of Hate Speech*, 2 AMST. L. F. 33 (2009); Richard Delgado & Jean Stefancic, *Four Observations About Hate Speech*, 44 WAKE FOREST L. REV. 353 (2009); Kathleen Mahoney, *Hate Speech, Equality, and the State of Canadian Law*, 44 WAKE FOREST L. REV. 321 (2009); ROBERT POST, *Hate Speech*, *in* EXTREME SPEECH AND DEMOCRACY (I. Hare & J. Weinstein eds., 2009); STEVEN P. LEE, *Hate Speech in the Marketplace of Ideas*, *in Freedom of Expression in a Diverse World* (D. Golash ed., 2010); Ishani Maitra & Mary Kate McGowan, *On Racist Hate Speech and the Scope of a Free Speech Principle*, 23 CAN. J. L. & JUR. 343 (2010); C. EDWIN BAKER, *Hate Speech*, *in* THE CONTENT AND CONTEXT OF HATE SPEECH: RETHINKING REGULATION AND RESPONSES (M. Herz & P. Molnar eds., 2012) [hereafter *Hate Speech*]; RONALD DWORKIN, *Reply to Jeremy Waldron*, *in* THE CONTENT AND CONTEXT OF HATE SPEECH: RETHINKING REGULATION AND RESPONSES (M. Herz & P. Molnar eds., 2012); ROBERT POST, *Interview*, *in* THE CONTENT AND CONTEXT OF HATE SPEECH: RETHINKING REGULATION AND RESPONSES (M. Herz & P. Molnar eds., 2012); NADINE STROSSEN, *Interview*, *in* THE CONTENT AND CONTEXT OF HATE SPEECH: RETHINKING REGULATION AND RESPONSES (M. Herz & P. Molnar eds., 2012); Jeremy Waldron, THE HARM IN HATE SPEECH (2012); CAROLINE WEST, *Words that Silence? Freedom of Expression and Racist Hate Speech*, *in* SPEECH & HARM (I. Maitra & M. K. McGowan eds., 2012); Kylie Weston-Scheuber, *Gender and the Prohibition of Hate Speech*, 12 QUEENSL. UNIV. TECH. L. & JUST. J. 132 (2012); Katharine Gelber & Luke J. McNamara, *Evidencing the Harms of Hate Speech*, 22 SOC. IDENTITIES 324 (2016); Eric Heinze, HATE SPEECH AND DEMOCRATIC CITIZENSHIP (2016); Alexander Brown, *Averting Your Eyes in the Information Age: Hate Speech, the Internet, and the Captive Audience Doctrine*, 12 CHARLESTON L. REV. 1 (2017); Alexander Brown, *Hate Speech Laws, Legitimacy, and Precaution: Reply to Weinstein*, 32 CONST. COMMENT. 599 (2017); James Weinstein, *Hate Speech Bans, Democracy, and Political Legitimacy*, 32 CONST. COMMENT. 527 (2017) [hereafter *Hate Speech Bans, Democracy, and Political Legitimacy*]; Alexander Brown, *Retheorizing Actionable Injuries in Civil Lawsuits Involving Targeted Hate Speech: Hate Speech as Degradation and Humiliation*, 9 ALA. C.R. & C.L. L. REV. 1 (2018); RICHARD DELGADO & JEAN STEFANCIC, MUST WE DEFEND NAZIS? HATE SPEECH, PORNOGRAPHY, AND THE NEW FIRST AMENDMENT (2018); CAROLINE WEST, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP (2018); ALEXANDER BROWN & ADRIANA SINCLAIR, THE POLITICS OF HATE SPEECH LAWS (2019) [hereafter THE POLITICS OF HATE SPEECH LAWS]; ALEXANDER TSESIS, FREE SPEECH IN THE BALANCE (2020); Suzanne Whitten, *A Recognition-Sensitive Phenomenology of Hate Speech*, 23 CRIT. REV. INT'L. SOC. & POL. PHIL. 853 (2020); Melina Constantine Bell, *John Stuart Mill's Harm Principle and Free Speech: Expanding the Notion of Harm*, 33 UTILITAS 162 (2021); Katharine Gelber, *Differentiating Hate Speech: A Systemic Discrimination Approach*, 24 CRIT. REV. INT'L. SOC. & POL. PHIL. 393 (2021); Gordan Ballingrud & Giovanna Scirrotto, Obscenity, *Hate Speech, and Viewpoint Discrimination: A Formula for Hate Speech as an Unprotected Category*, 20 DARTMOUTH L.J. 6 (2023); Samantha Barbas, *The Rise and Fall of Group Libel: The Forgotten Campaign for Hate Speech Laws*, 53 LOY. UNIV. CHI. L.J. 297 (2023);

Article has a more limited focus both because it concentrates only on the England and Wales stirring up hatred offenses and because it does not attempt to weigh up every argument for and against these offenses.

Second, in this Article I concentrate on the legal, as opposed to ordinary, concept of hate speech. The former concerns how legislatures and courts define illegal hate speech, such as in terms of the England and Wales stirring up hatred offenses, whereas the latter concerns how ordinary people use, and how major social institutions like media companies and Internet platforms define, the term "hate speech."[4] Nevertheless, I do not mean to suggest the two concepts exist entirely separately from each other. They can be interrelated in a variety of ways.[5] Indeed, legislators could potentially draft a law that effectively offers a definition of illegal stirring up of hatred that itself relies on the plain English or ordinary meaning of the term hate speech, such as along the following lines: "It is an offense for a person to use hate speech publicly and with intent thereby to stir up hatred against another person or group of people on grounds of protected characteristics." It would then be left to the jury to determine the ordinary meaning of the term hate speech using a reasonable person test. Perhaps legislators would be disinclined to draft such a law partly because they would consider it circular or because of the appearance of circularity. But, in fact, there is nothing circular about this formulation precisely because the legal concept of hate speech is distinct from the ordinary concept. Even though the term hate speech is being used in the above formulation in part to define a criminal offense, it is being interpreted as having its plain English or ordinary meaning. However, I do not advocate this formulation on the different grounds that it could be both viewpoint-discriminatory and overbroad. Instead, I propose and defend a model stirring up offense that retains the phrase "threatening words or behavior."

By way of an aside, the main focus of this Article is the England and Wales stirring up hatred offenses, but it is important to note that legislators in the United States would be unlikely to try to pass a law criminalizing the stirring up of hatred—or criminalizing group defamation—that was written so as to rely on the ordinary meaning of the term hate speech if

---

John Park, *The Mental and Physical Health Argument Against Hate Speech*, 9 J. COGNITION & NEUROETHICS 13 (2023).

4. For more on the distinction, *see* Alexander Brown, *What Is Hate Speech? Part 1: The Myth of Hate*, 36 L. & PHIL. 419, 421-61 (2017) [hereafter *What Is Hate Speech? Part 1*]; and ALEXANDER BROWN & ADRIANA SINCLAIR, HATE SPEECH FRONTIERS: EXPLORING THE LIMITS OF THE ORDINARY AND LEGAL CONCEPTS 4-12 (2023) [hereafter HATE SPEECH FRONTIERS].

5. BROWN & ADRIANA SINCLAIR, HATE SPEECH FRONTIERS, *supra* note 4, at ch. 5.

they believed that doing so would be futile, such as if they feared it would be struck down as unconstitutional by the U.S. Supreme Court because it considers hate speech a category of protected speech.[6] Then again, commercial speech is also a category of protected speech, only it has less protection. Could it be that hate speech is likewise a category of protected yet simultaneously less protected speech?[7] Then again, legislators on both sides of the pond might be further disinclined to try and pass any criminal law that relied on the ordinary meaning of the term hate speech on the grounds that this term is highly politicized, such that they would worry a jury could never convict a person of using hate speech to stir up hatred simply because individual jurors would be at loggerheads over the ordinary meaning of the term hate speech due to their own political and ideological differences. However, three points weigh against this fear. For one thing, even people who think the term hate speech is used by socially progressive elites to suppress speech they happen to dislike could still agree that the term has an ordinary meaning. In other words, they could agree on its ordinary meaning even if they disagreed strongly about its use as a pretext for censorship.[8] Furthermore, the England and Wales stirring up hatred offenses already contain the term "hatred," which, in itself, is undefined within the statute, meaning jurors are expected to interpret it as having its plain English or ordinary meaning. Yet, nonetheless, jurors have not found it difficult to reach agreement and convict guilty persons. Lastly, in the U.K., under the so-called golden rule of statutory interpretation, courts have an obligation to follow the ordinary meanings of undefined statutory terms unless and until doing so would run contrary to the legislative purpose of those very statutes, for example.[9] Thus, if some jurors were minded to interpret the ordinary meaning of the term hate speech to mean "any speech disliked by the PC brigade but which manifestly ought never to be prohibited," then this would clearly cut against the intentions of parliamentarians in creating the stirring up hatred offenses, and the judge could instruct those jurors not to adopt such a reading. All that being said, I so not defend the aforementioned legislative approach in this Article, since my aim is to compare the England and Wales stirring up offenses against two particular aspects of First Amendment jurisprudence, namely, viewpoint discrimination and

---

6. *See infra* note 24.

7. For a full defense of the reading that hate speech could be a category of protected yet less protected speech, *see* Alexander Brown, *"Bigots in Black Robes": Legal Ethics and Judicial Hate Speech*, 20 INTERCULTURAL HUM. RTS. L. REV. (2025).

8. *See* Brown, *What Is Hate Speech? Part 1*, *supra* note 4, at 425-26.

9. Grey v. Pearson, 10 E.R. 1216 (1857).

overbreadth, and because I believe that together these aspects favor retaining the statutory phrase "threatening words or behavior."

Third, as stated above, this Article assesses what it would take to avoid or mitigate unjustifiable viewpoint discrimination in the England and Wales stirring up hatred offenses, and what this could mean in terms of increasing the risk of overbreadth. (Of course, by discussing two First Amendment doctrines, I am also consciously placing this Article within the discipline of comparative law.) To explain, in *Rosenberger v. Rector and Visitors of University of Virginia*,[10] the U.S. Supreme Court held that viewpoint discrimination is an "egregious form of content discrimination" and is "presumptively unconstitutional,"[11] thus lending it a high status within First Amendment jurisprudence. Indeed, some scholars have labeled viewpoint discrimination "a cardinal sin under the First Amendment."[12] One way to mitigate against viewpoint discrimination is to reform a law to make it wider in scope, so that it no longer favors one viewpoint over another. However, making laws wider in scope can increase the risk of them being overbroad. Roughly speaking, a law is overbroad if, as well as prohibiting unprotected speech, it also sweeps up a substantially disproportionate amount of constitutionally protected speech. As described in *Broadrick v. Oklahoma*,[13] for example, when courts apply the overbreadth doctrine it can sometimes mean that "any enforcement of a statute thus placed at issue is totally forbidden until and unless a limiting construction or partial invalidation so narrows it as to remove the seeming threat or deterrence to constitutionally protected expression."[14] Viewed in such terms, "[a]pplication of the overbreadth doctrine in this manner is, manifestly, strong medicine."[15]

Therefore, in practice, the aim of avoiding or mitigating viewpoint discrimination sits in tension with the requirement of avoiding overbreadth. Even if a law did *not* involve unjustifiable viewpoint discrimination—or even if legislators redrafted a law to avoid or reduce viewpoint discrimination—it might, as a result, prove to be unconstitutionally overbroad. This is precisely what Justice White opined in *R.A.V. v. City of St. Paul*.[16] Whilst disagreeing with Justice Scalia's

---

10. Rosenberger v. Rector and Visitors of University of Virginia, 515 U.S. 819, 829 (1995).

11. *Id.* at 829-30.

12. Clare R. Norins & Mark L. Bailey, *Campbell v. Reisch: The Dangers of the Campaign Loophole in Social-Media-Blocking Litigation*, 25 J. CONST. L. 146, 147 (2023).

13. Broadrick v. Oklahoma, 413 U.S. 601, 613 (1973).

14. *Id.* at 613.

15. *Id.*

16. R.A.V. v. City of St. Paul, 505 U.S. 377, 381-82 (1992).

assessment that the city's cross-burning ordinance was unjustifiably viewpoint discriminatory, Justice White nonetheless judged the ordinance to be overbroad.[17] These two requirements demand careful calibration, if not equilibrium.

Interestingly, in recent years the U.S. Supreme Court has sought to rebalance or lessen the weight and leeway given to the overbreadth doctrine. For example, in *U.S. v. Hansen*,[18] the Court upheld a conviction under a U.S. Code which forbids "encourag[ing] or induc[ing] an alien to come to, enter, or reside in the United States, knowing or in reckless disregard of the fact that such [activity] is or will be in violation of law." The Court held that "[t]o justify facial invalidation, a law's unconstitutional applications must be realistic, not fanciful, and their number must be substantially disproportionate to the statute's lawful sweep."[19] Arguably, the Court qualified the overbreadth doctrine in this way in order to reduce the judicial pressure (perverse incentive) on legislators to make laws narrower than would be optimally effective in combating the very speech or conduct that the laws are designed to combat. (Qualifying the overbreadth doctrine also chimes with a broader tradition of jurisprudence that makes a virtue out of judicial restraint: This is the idea that a good judge does not seek to "legislate from the bench" and that good courts do not act as "roving commissions assigned to pass judgment on the validity of the nation's laws.")

However, all of this begs the question: Why does it matter if the England and Wales stirring up hatred offenses involve viewpoint discrimination and/or overbreadth? Surely it is otiose or "academic" whether non-U.S. hate speech laws would pass muster under U.S. First Amendment jurisprudence given that jurisdictional boundaries mean there is no constitutional reason, from a British perspective, why the former must pass muster under the latter. I believe there are three substantive reasons it matters, over and above the brute fact that there is a long tradition of legal scholarship comparing non-U.S. hate speech laws against First Amendment doctrines.[20] First, given the rise of judicial

17. *Id.* at 408-11.

18. U.S. v. Hansen, 599 U.S. 5 (2023).

19. *Id.*

20. Heinze, *Viewpoint Absolutism and Hate Speech*, *supra* note 3; WEINSTEIN, *Hate Speech, Viewpoint Neutrality, and the American Concept of Democracy*, *supra* note 3; Kevin Boyle, *Hate Speech—The United States Versus the Rest of the World*, 53 ME. L. REV. 488 (2001); EDWARD J. EBERLE, DIGNITY AND LIBERTY: CONSTITUTIONAL VISIONS IN GERMANY AND THE UNITED STATES (2002); FREDERICK SCHAUER, *Freedom of Expression Adjudication in Europe and America: A Case Study in Comparative Constitutional Architecture*, in EUROPEAN AND U.S. CONSTITUTIONALISM (G. Nolte ed., 2005); ADRIENNE STONE, *How to Think About the Problem of*

globalization, and the cross-pollination of legal ideas and norms relating not merely to free speech but also to hate speech among and between judges from different parts of the world,[21] analyzing whether non-U.S. hate speech laws would survive certain aspects of the U.S. Supreme Court's strict scrutiny test is not idle speculation. On the contrary, it may be something that U.K. Supreme Court judges—and judges of the European Court of Human Rights (ECtHR), to which U.K. cases can be appealed—consider in their decision-making. Second, even if U.K. Supreme Court judges—and ECtHR judges—are not influenced by what happens across the pond, maybe they ought to be. Perhaps there are moral reasons why non-U.S. hate speech laws should pass muster under U.S. First Amendment jurisprudence. For, against the backdrop of a range of normative arguments for and against hate speech laws, it could conceivably tip the scales one way or the other (at least in some people's eyes) if non-U.S. hate speech laws could pass muster under U.S. First Amendment jurisprudence, given the many important and universal values this jurisprudence invokes in support of free speech. Third, international cooperation in combating hate speech could depend, to some small extent, on extradition and mutual legal assistance and these may be reliant upon requirements of double criminality.[22]

Of course, there is much more to the U.S. Supreme Court's approach to the First Amendment than the doctrines of viewpoint discrimination and overbreadth. For one thing, the Court has, in part, adopted a categorical approach to free speech, distinguishing between protected and

---

Hate Speech: Understanding a Comparative Debate, in HATE SPEECH AND FREEDOM OF SPEECH IN AUSTRALIA (K. Gelber & A. Stone eds., 2007); PETER MOLNAR, *Towards Improved Law and Policy on "Hate Speech": The "Clear and Present Danger" Test in Hungary*, *in* EXTREME SPEECH AND DEMOCRACY (I. Hare & J. Weinstein eds., 2010); ERIK BLEICH, THE FREEDOM TO BE RACIST? HOW THE UNITED STATES AND EUROPE STRUGGLE TO PRESERVE FREEDOM AND COMBAT RACISM (2011); Baker, *Hate Speech*, *supra* note 3; EDUARDO BERTONI & JULIO RIVERA, *The American Convention on Human Rights: Regulations on Hate Speech and Other Similar Expressions*, *in* THE CONTENT AND CONTEXT OF HATE SPEECH: RETHINKING REGULATION AND RESPONSES (M. Herz & P. Molnar eds., 2012); Weinstein, *Hate Speech Bans, Democracy, and Political Legitimacy*, *supra* note 3; Robert M. O'Neil, *Hate Speech, Fighting Words, and Beyond—Why American Law Is Unique*, 76 ALBANY L. REV. 467 (2013).

21. See Anne-Marie Slaughter, *Judicial Globalization*, 40 VA. J. INT'L L. 1103 (2000); ANNE-MARIE SLAUGHTER, *A Brave New Judicial World*, in *American Exceptionalism and Human Rights* (M. Ignatieff ed., 2005); and BROWN & SINCLAIR, HATE SPEECH FRONTIERS, *supra* note 4, at ch. 7.

22. *See also* OCTOPUS PROJECT & ALEXANDER BROWN, IMPLEMENTING THE FIRST PROTOCOL TO THE CONVENTION ON CYBERCRIME ON XENOPHOBIA AND RACISM: GOOD PRACTICE STUDY 46 (2023).

unprotected categories of speech,[23] and, importantly, it has repeatedly identified hate speech as a category of protected speech.[24] The fact that hate speech is protected speech brings a core principle of First Amendment jurisprudence directly into play, namely that "government has no power to restrict expression because of its message, its ideas, its subject matter, or its content" (the rule against content discrimination).[25] Other relevant features of First Amendment jurisprudence include the fact that the test for proscribable incitement is inciting "imminent lawless action,"[26] while the test for proscribable threats is "true threats."[27] It is highly likely that the mere use of threatening words or behavior with intent to stir up racial hatred, for example, falls short of both these doctrinal benchmarks. However, the mere fact that the viewpoint discrimination and overbreadth doctrines do not all pass muster under First Amendment jurisprudence across the board, does not mean that these doctrines are not worth looking into on their own merits and in isolation from other doctrines. There are important values served by these doctrines that would be better served by upholding them than by ignoring them, irrespective of the remainder of First Amendment jurisprudence.

Finally, I need to make clear that I am relying on a particular concept of autonomy. In this Article I say relatively little about what C. Edwin Baker calls "formal autonomy," which has to do with the state's respect for people's freedom to decide what to say and receive.[28] Substantive autonomy, by contrast, has to do with arranging the total framework of freedoms and restrictions so as to ensure that one person's choice of words or behavior does not diminish the autonomy of another person, such as by going around or cutting short their independent, rational deliberations.[29] Baker himself cites several examples of what he *does* consider to be coercive speech of the sort that threatens substantive autonomy, or

---

23. *See, e.g.*, U.S. v. Stevens, 559 U.S. 460 (2010).

24. This is evidenced in the following opinions handed down by the U.S. Court of Appeals and the U.S. Supreme Court respectively. "One of the things that separates our society from [societies ruled by totalitarian governments] is our absolute right to propagate opinions that the government finds wrong or even hateful." American Booksellers Association v. Hudnut, 771 F.2d 323, 327-8 (U.S. Court of Appeals, 7th Cir., 1985). "[Speech] cannot be restricted simply because it is upsetting or arouses contempt." Snyder v. Phelps, 562 U.S. 443, 458 (2011). "Speech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful; but the proudest boast of our free speech jurisprudence is that we protect the freedom to express 'the thought that we hate.'" Matal v. Tam, 582 U.S. 218 (2017).

25. Police Department of Chicago v. Mosley, 408 U.S. 92, 95 (1972).

26. Brandenberg v. Ohio, 395 U.S. 444, 447 (1969).

27. Watts v. United States 394 U.S. 705, 708 (1969).

28. Baker, *Autonomy and Hate Speech*, *supra* note 3 at 142.

29. *Id.* at 143.

"speech designed to disrespect and distort the integrity of another's mental processes or autonomy,"[30] namely speech used in the enactment of fraud, perjury, blackmail, espionage, and treason.[31] Other writers allow similar exceptions. David A. Strauss, for example, cites "two categories of speech that move people to action by means other than the rational process of persuasion: namely, false statements and speech that seeks to elicit action before the hearer has thought about the speech and possible answering arguments."[32] Importantly, there exists another tradition of academic research that places certain kinds of hate speech (e.g. slurs, group defamation, negative stereotypes, certain provocations) into the class of speech that threatens substantive autonomy, such as by circumventing or curtailing receivers' independent, rational deliberations.[33] This Article both falls squarely within and advances this tradition. In a giant leap forward for hate speech law, it draws on a range of insights about how stirring up can threaten receivers' autonomy to recommend quite significant reforms to the England and Wales stirring up hatred offenses.

II. THE SOCIAL CONTEXT AND CONTENT OF THE ENGLAND AND WALES STIRRING UP HATRED OFFENSES

Tommy Robinson is a far-right political leader and activist who, over two decades, has occupied various positions of trust, authority, and influence in British political affairs, the large and the small.[34] He has repeatedly faced allegations of using Islamophobic hate speech. For example, in 2011, he gave a speech at a London demonstration that was video recorded and posted to YouTube in which he declared:

---

30. C. EDWIN BAKER, *The Liberty Theory*, *in* HUMAN LIBERTY AND FREEDOM OF SPEECH 59-60 (1989).

31. *Id.* at 60.

32. David A. Strauss, *Persuasion, Autonomy, and Freedom of Expression*, 91 COL. L. REV. 334, 335-336 (1991).

33. *See* Susan Brison, *The Autonomy Defense of Free Speech*, 108 ETHICS 312, 328 (1998); David Brink, *Millian Principles, Freedom of Expression, and Hate Speech*, 7 LEG. THEORY 119, 138-140 (2001); Andres Moles, *Autonomy, Free Speech and Automatic Behaviour*, 13 RES PUBLICA 53 (2007); Brown, HATE SPEECH LAW, *supra* note 2, at 62-66.

34. For example, he has served as joint vice-chairman of the far-right British Freedom Party (BFP), as cofounder and leader of the far-right English Defence League (EDL), as political advisor to Gerard Batten, leader of the U.K. Independence Party (UKIP) (a mainstream British political party that previously played an important role in Brexit), and as an online influencer on controversial issues relating to immigration, Muslim relations in Britain, and Russia's invasion of Ukraine.

> Every single Muslim watching this on YouTube, on 7/7 you got away with killing and maiming British citizens, you got away with it. You had better understand that we have built a network from one end of this country to the other end, and we will not tolerate it, and the Islamic community will feel the full force of the English Defence League if we see any of our citizens killed, maimed or hurt on British soil ever again.[35]

In 2023, Robinson publicly thanked Elon Musk for reinstating his X account,[36] which now has nearly one million followers. A few months later, Robinson used his profile to post numerous comments and videos in the wake of the fatal mass stabbings of children and adults at a Taylor Swift-themed holiday club in Southport at the end of July 2024. Musk himself engaged with one of Robinson's posts—that had criticized the U.K. prime minister for (in Robinson's eyes) handling the public protests following the Southport stabbings with an uneven and heavy hand[37]—by replying with double exclamation marks.[38] For its part, the charity Hope Not Hate alleged that, despite being located outside of Britain, Robinson had successfully used his reinstated X account to stoke up Islamophobia in response to the Southport stabbings before any facts had been officially established concerning the identity of the suspect.[39] Here are three of Robinson's posts from the days following the stabbing:

> There's more evidence to suggest islam is a mental health issue rather than a religion of peace.[40]

> 3 children were brutally murdered at a Taylor Swift dance class yesterday and the police are "managing" the situation in order to keep the people quiet! Hundreds of Muslims hit Rochdale police station because one of them attacked officers and got a kick, and police didn't even intervene. The people of Southport are rightfully angry, and I don't blame them one bit.[41]

> Mobs of Muslims have been running around numerous towns all day attacking people protesting following the horrific murders of 3 children at

35. Natasha Red, *Tommy Robinson (EDL) Threatens "Every Single Muslim" in the UK*, YOUTUBE (Sept. 7, 2011), https://www.youtube.com/watch?v=8j7IX_5a_9M.

36. @TRobinsonNewEra, X (Nov. 5, 2023).

37. @TRobinsonNewEra, X (Aug. 1, 2024).

38. @elonmusk, X (Aug. 1, 2024).

39. Nadine White, *Tommy Robinson Stokes Far-Right Riots on Social Media from Outside UK*, THE INDEPENDENT (Aug. 6, 2024, 5:43 AM), https://www.independent.co.uk/news/uk/home-news/tommy-robinson-uk-riots-edl-twitter-b2591161.html.

40. @TRobinsonNewEra, X (July 30, 2024).

41. @TRobinsonNewEra, X (July 30, 2024).

> a Taylor Swift dance class after being told by @Keir_Starmer and the media they're "far right edl thugs" and the police have done nothing.[42]

However, to date, Robinson has faced no criminal prosecutions for alleged crimes committed under the England and Wales stirring up hatred offenses. (That being said, he has previously been found guilty of contempt of court for posting videos of himself outside of court in other people's cases.[43])

There are a few likely reasons for the lack of prosecutions against Robinson under the stirring up hatred offenses. A jurisdictional reason is that, in the cases of the X posts from July and August 2024, Robinson made the posts whilst outside of the U.K. (ironically whilst "on the run" from yet further contempt of court charges). Therefore, my first recommendation is that the stirring up hatred offenses be given extraterritorial applicability, meaning they should be made applicable even to acts done outside of England and Wales provided they are done by a person who is habitually resident in England and Wales.

Another reason is due to how narrowly the offenses are drawn up. For example, when it comes to offenses involving the stirring up religious hatred set out in Part 3A of the Public Order Act 1986, the prosecution must show that the defendant not only used "threatening words or behaviour" but also did so with intent thereby to "stir up religious hatred." Arguably, Robinson's speech and YouTube video from 2011 amounted to threatening words. But perhaps the police and prosecutors were uncertain that they could prove beyond a reasonable doubt that Robinson had used these words with intent to stir up religious hatred specifically. I make some further recommendations to deal with this issue in Part III.

Importantly, Hope Not Hate also argued that Robinson's posts contributed to a climate of hatred, fear, and anger that was itself partly responsible for the widespread violent protests, riots, mass brawls, and assaults in Britain following the Southport stabbings, typically involving young male followers of far-right political movements and young male members of the British Muslim community.[44] These most recent public disturbances, to some degree, echo race riots of the past in Britain. Such riots have frequently been cited by parliamentarians as a rationale for

---

42. @TRobinsonNewEra, X (Aug. 3, 2024).

43. Attorney General's Office, Press Release: Stephen Yaxley-Lennon Committed to Prison for Contempt of Court (July 11, 2019), https://www.gov.uk/government/news/stephen-yaxley-lennon-committed-to-prison-for-contempt-of-court.

44. White, *supra* note 39.

introducing the stirring up hatred offenses in the first place.[45] However, critics of such laws have also argued that maintaining public order could merely serve as a pretext for any governments who are in fact seeking to ban ideas or opinions they simply find reprehensible or even perhaps politically inconvenient.[46] Indeed, the fact that the England and Wales stirring up hatred offenses do not contain, as an essential element, that the relevant words caused or were likely to cause a breach of the peace, means that, according to critics, the police are being given powers to act as "censors of speech."[47] In other words, such laws amount to the state taking sides in public disputes over matters of opinion, which it ought not do.[48] Indeed, some parliamentarians have argued that these laws can actually worsen social tensions, such as if the public feels that certain minority groups are receiving special protections at the expense of the freedoms of so-called "ordinary people."[49] It is certainly arguable that these laws pose a threat to people's formal autonomy insofar as they involve the state, rather than individuals, deciding what ideas or opinions they will give and receive.

In fact, the England and Wales stirring up hatred offenses are a mixed collection of offenses, as set out in Parts 3 and 3A of the Public Order Act 1986 (with jurisdiction in England and Wales), Part 3 of the Public Order (Northern Ireland) Order 1987 (with jurisdiction in Northern Ireland), and Part 3 of the Hate Crime and Public Order Act 2021 (with jurisdiction in Scotland). In what follows, I shall solely focus on the versions of the offenses found in England and Wales, unless otherwise stated.[50] Within this jurisdiction, the relevant offenses cover "words or behaviour" but also the "displaying of written material"; some offenses relate to "threatening, abusive or insulting words or behaviour," whilst others refer only to "threatening words or behaviour"; some, but not all, offenses contain sections explicitly aimed at the "[p]rotection of freedom of expression"; the offenses include within their scope some, but not all, kinds of legally recognized protected characteristics; and they apply substantially different legal tests or elements to different protected characteristics.[51] For example, according to Section 18(1),

---

45. *See* BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3, AT CH. 3; AND JEN NELLER, STIRRING UP HATRED: MYTH, IDENTITY AND ORDER IN THE REGULATION OF HATE SPEECH (2023).

46. BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3.

47. Neller, *supra* note 45, at 127.

48. Heinze, *Viewpoint Absolutism and Hate Speech*, *supra* note 3.

49. Neller, *supra* note 45, at 216.

50. Public Order Act 1986, Parts 3 and 3A.

51. *Id.*

> A person who uses threatening, abusive or insulting words or behaviour, or displays any written material which is threatening, abusive or insulting, is guilty of an offence if (a) he intends thereby to stir up racial hatred, or (b) having regard to all the circumstances racial hatred is likely to be stirred up thereby.[52]

By contrast, according to Section 29B(1), "A person who uses threatening words or behaviour, or displays any written material which is threatening, is guilty of an offence if he intends thereby to stir up religious hatred or hatred on the grounds of sexual orientation."[53] Thus, in its 2021 report on extending and reforming the stirring up hatred offenses, the Law Commission of England and Wales recommended both the inclusion of sex or gender and disability as protected characteristics and the standardization of the legal requirements (or elements) of the offenses across the different protected characteristics.[54]

It is also important to note that, under Sections 27(3) and 29L of the Public Order Act 1986, the England and Wales stirring up hatred offenses are *either-way* offenses.[55] This means that less serious cases of stirring up hatred can be charged as summary offenses with lower maximum sentences and decided by magistrates without a jury, while more serious cases can be charged as indictable offenses with much higher maximum sentences and decided by a jury under the guidance of a judge or recorder in crown court.

III. VIEWPOINT DISCRIMINATION AND OVERBREADTH

Building on the core principle that the government ought not to engage in content discrimination, the U.S. Supreme Court has identified viewpoint discrimination as being an especially problematic form of content discrimination.[56] In its words:

> [T]he First Amendment forbids the government to regulate speech in ways that favor some viewpoints or ideas at the expense of others.[57]

---

52. *Id.* at § 18(1).

53. *Id.* at § 29B(1).

54. LAW COMMISSION, HATE CRIME LAWS: FINAL REPORT, LAW COM. NO. 402 (2021), Recomms. 21, 22 and 23, at 541.

55. Public Order Act 1986, at §§ 27(3) and 29L.

56. Rosenberger v. Rector and Visitors of University of Virginia, 515 U.S. 819, 829 (1995).

57. City Council of Los Angeles v. Taxpayers for Vincent, 466 U.S. 789, 804 (1984).

> [T]he government violates the First Amendment when it denies access to a speaker solely to suppress the point of view he espouses on an otherwise includible subject.[58]
>
> The government may not regulate use based on hostility—or favoritism—towards the underlying message expressed.[59]
>
> The government must abstain from regulating speech when the specific motivating ideology or the opinion or perspective of the speaker is the rationale for the restriction.[60]
>
> [S]peech discussing otherwise permissible subjects cannot be excluded [. . .] on the ground that the subject is discussed from a religious viewpoint.[61]
>
> [The disparagement clause of the Langham Act] denies registration to any mark that is offensive to a substantial percentage of the members of any group. But in the sense relevant here, that is viewpoint discrimination: Giving offense is a viewpoint.[62]
>
> [T]his Court invalidated the Lanham Act's bar on the registration of "disparag[ing]" trademarks [. . .] because it discriminated on the basis of viewpoint. Today we consider a First Amendment challenge to [. . .] prohibiting the registration of "immoral [. . .] or scandalous" trademarks. We hold that this provision infringes the First Amendment for the same reason: It too disfavors certain ideas.[63]

Why is viewpoint discrimination held to be an especially problematic form of content discrimination? Suppose one believes that one of the fundamental values served by the First Amendment is the right of citizens to participate as equals in the formation of public opinion upon which political decision-making is based.[64] For the state to treat citizens as having this right entails that the state should leave it to citizens to decide for themselves what opinions shall be expressed. In the words of the U.S. Supreme Court again:

> The constitutional right of free expression is powerful medicine in a society as diverse and populous as ours. It is designed and intended to remove

58. Cornelius v. NAACP Legal Defense and Educational Fund, 473 U.S. 788, 806 (1985).
59. R.A.V. v. City of St. Paul, at 386.
60. Rosenberger v. Rector and Visitors of University of Virginia, at 829.
61. Good News Club v. Milford Central School, 533 U.S. 98, 112 (2001).
62. Matal v. Tam, 582 U.S. 218 (2017).
63. Iancu v. Brunetti, 588 U.S. 388 (2019).
64. Weinstein, *Hate Speech Bans, Democracy, and Political Legitimacy*, *supra* note 3, at 528.

> governmental restraints from the arena of public discussion, putting the decision as to what views shall be voiced largely into the hands of each of us, in the hope that use of such freedom will ultimately produce a more capable citizenry and more perfect polity [. . .].[65]

This also entails that the state should not seek to pick and choose who may participate simply based on what the state thinks about certain viewpoints. As James Weinstein has pointed out, if the state were to restrict some viewpoints and not others based on which it favors, then public opinion "would reflect the will of the governing officials rather than the will of the people."[66] Therefore, the state should not take sides in public debates, such as by suppressing some viewpoints but not others, even if the state finds certain viewpoints to be particularly disagreeable, distasteful, immoral, or hateful.

To clarify, I am not suggesting that defending the rule against viewpoint discrimination is only possible by appealing to democracy-based arguments. As Richard H. Fallon puts it, "virtually all of the leading theories would find it impermissible—albeit for different reasons—for the government to attempt to stifle communication based on its hostility to particular ideas."[67] Take theories of the First Amendment that highlight its role in protecting formal autonomy. If the argument from democracy places the emphasis on decisions about which viewpoints can be aired reflecting not the will of government, but the will of the people as a collective entity, then the argument from formal autonomy shifts the focus to the will of the individual. On this view, viewpoint discrimination infringes the autonomy of individuals in deciding for themselves about given speech as opposed to having that decision made on their behalf by the state, including both the speaker's autonomy in deciding whether to say things others may find disagreeable and the receiver's autonomy in deciding whether they find other people's speech disagreeable.[68]

At any rate, there is a common assumption in the free speech literature that the England and Wales stirring up hatred offenses, and similar laws elsewhere in the world banning either incitement to hatred or group defamation, involve viewpoint discrimination of the sort that would be unlikely to pass the strict scrutiny test of U.S. First Amendment

---

65. *Cohen v. California*, 403 U.S. 15, 24 (1971).

66. Weinstein, *Hate Speech, Viewpoint Neutrality, and the American Concept of Democracy*, *supra* note 3, at 147.

67. Richard H. Fallon, *Implementing the Constitution*, 111 HARV. L. REV. 54, 100 (1997).

68. *See* C. Edwin Baker, *The Independent Significance of the Press Clause Under Existing Law*, 35 HOF. L. REV. 955, 980 (2007); *and* C. Edwin Baker, *Autonomy and Free Speech*, 27 CONST. COMMENT. 251, 254 (2011).

jurisprudence if, contrary to fact, there were a legal requirement upon non-U.S. hate speech laws to pass this test.[69] Having assumed the England and Wales stirring up hatred offenses do involve viewpoint discrimination, scholars shift the thought experiment into another gear by asking whether these offenses, and other hate speech laws, could nevertheless be warranted or justifiable on some valid basis or legitimate ground.[70] The latter discussion has often included asking whether a case could be made for saying the offenses, and other hate speech laws, could fall under one of the exceptions permitting content and viewpoint discrimination envisaged by the U.S. Supreme Court in *R.A.V.*[71] (This discussion deliberately puts to one side the fact that this decision concerned exceptions for laws banning a subcategory of *unprotected* speech whereas hate speech has been identified as a category of *protected* speech within First Amendment jurisprudence.[72]) Nevertheless, the aim of this part is to critically assess both the initial common assumption that the England and Wales stirring up hatred offenses involve viewpoint discrimination and the further argument about whether any viewpoint justification could be justified.

Of course, having said this is a common assumption, it is also fair to point out there are exceptions. So, for example, the legal scholar Philip

---

69. *See, e.g.*, Michael A. G. Korengold, *Lessons in Confronting Racist Speech: Good Intentions, Bad Results, and Article 4(a) of the Convention on the Elimination of All Forms of Racial Discrimination*, 77 MINN. L. REV. 719, 727 (1993); Susannah C. Vance, *The Permissibility of Incitement to Religious Hatred Offenses under European Convention Principles'*, 14 TRANSNAT'L L. & CONTEMP. PROBLS. 201, 214 (2004); HARE, *supra* note 1, at 531; HEINZE, VIEWPOINT ABSOLUTISM AND HATE SPEECH, *supra* note 3, at 547-48; Roger Kiska, *Hate Speech: A Comparison Between the European Court of Human Rights and the United States Supreme Court Jurisprudence*, 25 REGENT. UNIV. L. REV. 107, 147-50 (2013); WEINSTEIN, HATE SPEECH BANS, DEMOCRACY, AND POLITICAL LEGITIMACY, *supra* note 3, at 545; Lackland H. Bloom, *The Rise of the Viewpoint-Discrimination Principle*, 72 SMU L. REV. F. 20, 39 (2019); Angelo Ryu, *Hate Speech and Public Reason*, OXFORD UNIV. UNDERGRADUATE L.J. 217, 234 (2020).

70. *See* MICHEL ROSENFELD, *Hate Speech in Constitutional Jurisprudence: A Comparative Analysis*, *in* THE CONTENT AND CONTEXT OF HATE SPEECH: RETHINKING REGULATION AND RESPONSES 284-85 (M. Herz & P. Molnar eds., 2012); Gideon Elford, *Legitimacy, Hate Speech, and Viewpoint Discrimination*, 18 J. MORAL PHIL. 239 (2021); Sebastien Bishop, *Should We Hate Hate Speech Regulation? The Argument from Viewpoint Discrimination*, 74 PHIL. Q. 1059 (2024).

71. *See* Heidi Kitrosser, *Containing Unprotected Speech*, 57 FLA. L. REV. 843 (2005); STEVEN J. HEYMAN, *Hate Speech, Public Discourse and the First Amendment*, *in* EXTREME SPEECH AND DEMOCRACY 164n.27 (I. Hare & J. Weinstein eds., 2009); Corey Brettschneider, *Value Democracy as the Basis for Viewpoint Neutrality: A Theory of Free Speech and Its Implications for the State Speech and Limited Public Forum Doctrines*, 107 NW. UNIV. L. REV. 603, 607 (2013); BROWN, HATE SPEECH LAW, *supra* note 2, at 287-97.

72. *See* American Booksellers Ass'n, Inc. v. Hudnut, 711 F.2d 323, 327-28 (7th Cir. 1985); Snyder v. Phelps, 562 U.S. 443, 458 (2011); Matal v. Tam, 582 U.S. 218 (2017).

N.S. Rumney has previously sought to challenge the assumption that the England and Wales stirring up hatred offenses are viewpoint-discriminatory by pointing to the fact that the legislation covers protected characteristics such as race and ethnicity rather than particular groups such as Black people, and, moreover, by citing a case, *R. v. Malik*,[73] in which the offenses were applied to a Black speaker who allegedly stirred up racial hatred against whites.[74] According to Rumney, the fact that the legislation is multidirectional, both as written and as applied by courts, shows that,

> it draws the line at any speech that incites racial hatred. In other words, particular viewpoints are not outlawed. Rather, it is the manner in which the words are communicated that is regulated.[75]

However, whilst it is certainly true that hate speech laws can potentially involve viewpoint discrimination, whether as written or as applied or both, Rumney's attempt to exonerate the England and Wales stirring up hatred offenses of being viewpoint-discriminatory falls flat upon closer inspection. Granted, the offenses do not permit using threatening words or behavior to stir up hatred against Black or white people on grounds of race or ethnicity, they still do permit using threatening words to stir up hatred against men, women, transpeople, gender-fluid people, and agender or gender-free people on the grounds of gender identity, for example. This too smacks of viewpoint discrimination, because the offenses do not permit certain viewpoints on race and ethnicity, while permitting viewpoints on gender identity.

Indeed, the situation is even more complicated in Scotland. Here the legislation creates several subcategories of stirring up hatred offenses relating to the protected characteristics of race, color, nationality (including citizenship), or ethnic or national origins, age, disability, religion or, in the case of a social or cultural group, perceived religious affiliation, sexual orientation, transgender identity, and variations in sex characteristics.[76] But given the omission of the broader characteristic of "gender identity," this means that it would not be permitted to stir up hatred against Black people or transgender people, for example, but would be permitted to stir up hatred against women.

---

73. R. v. Malik (1968) 140 (Lord Justices Winn and James of England) Cr. App. R(S).

74. Philip N.S. Rumney, *The British Experience of Racist Hate Speech Regulation: A Lesson for First Amendment Absolutists?*, 32 COMMON L. WORLD REV. 117 (2003).

75. *Id.* at 152.

76. Part 3 of the Hate Crime and Public Order (Scotland) Act 2021.

Therefore, in order to more robustly test the claim that the England and Wales stirring up hatred offenses, and similar hate speech laws, are unjustifiably viewpoint-discriminatory, I examine four additional pairwise comparisons of impermissible and permissible conduct under the legislation.

The first pairwise comparison reflects a key feature of the England and Wales stirring up hatred offenses: that they focus solely on the stirring up of *hatred* as opposed to the stirring up of a wide range of things. As made clear by the Explanatory Notes to the Racial and Religious Hatred Bill—a bill designed to extend the scope of the offenses to cover the stirring up of religious as well as racial hatred: "[t]he offences will not encompass material that just stirs up ridicule or prejudice or causes offence."[77] Thus, consider the following pair of examples.

Pairwise Comparison 1:

Not permitted: A person who uses threatening words or behavior, publicly, with intent to stir up hatred against Black people.

e.g. [words] "Black people are involved in a conspiracy to replace white people and so deserve our hatred; we should be prepared to defend white people with violence if necessary and the fact that they force us into this position is yet another reason to hate them."

e.g. [behavior] A person sees some people waiting to board a bus he is travelling on. He looks at everyone on the bus, and then walks to the entrance of the bus and allows a white person to get on but then blocks the way of a Black person, standing with his fists clenched, staring down at the person and then staring at the other passengers and shaking his head.

Permitted: A person who uses threatening words or behavior, publicly, with intent to stir up prejudice against Black people.

e.g. [words] "Black people are so god damn lazy, the only way to get them to work is by threats of violence; slave owners knew how to treat their [slaves], we should take leaf out of their book."

e.g. [behavior] A group of people perform a play wearing blackface and depicting Black people as idle, drunk, smiling fools who are cajoled into work by slave owners using horse whips.

This first pairwise comparison might be thought to reveal viewpoint discrimination because the legislation has the effect of restricting a viewpoint about why Black people deserve to be hated but permitting a

77. Explanatory Notes to the Racial and Religious Hatred Bill [Bill 11-EN] para. 15 (June 9, 2005), https://publications.parliament.uk/pa/cm200506/cmbills/011/en/06011x—.htm.

viewpoint about Black people being stereotypically lazy. As the U.S. Supreme Court puts it: "the First Amendment forbids the government to regulate speech in ways that favor some viewpoints or ideas at the expense of others."[78]

One response to this accusation of viewpoint discrimination is to accept the force of the argument and, therefore, redraft the stirring up hatred offense to encompass a wider swath of negative emotions, sentiments, or attitudes that a person is not permitted to stir up. If the rationale for the law has something to do with the social ills that can result from the stirring up of hatred, then why not extend the law so that it captures the stirring up of other negative emotions, sentiments, or attitudes, such as prejudice, intolerance, contempt, resentment, and anger, for instance? Maybe the ordinary meaning of the term hatred comes close to these other things, yet it might not be synonymous with them in the minds of jurors. Statutory inclusion of other terms besides hatred could widen the law to capture more directly and less ambiguously the stirring up of these other things.

No doubt some people would argue this response goes too far in mitigating the risk of viewpoint discrimination, that it opens the door to rendering the offenses overbroad. It might be difficult to say anything negative about a particular person or group of people identified by a protected characteristic without running the risk of being seen to promote negative emotions, sentiments, or attitudes about those people. For example, could a person acknowledge their own unconscious bias towards Black people without leaving themselves open to prosecution?

One counter to this concern is to say that any reforms to the England and Wales stirring up hatred offenses should retain the element of intent. Additionally, reforms should pay attention to the core rationales for introducing the offenses in the first place.[79] One important rationale was to combat the indirect threat to public order that can be created by the stirring up of hatred—for example, an indirect threat of violent protests, mass brawls, riots, and assaults of the sort that have occurred sporadically in Britain for centuries including most recently in the wake of the 2024 Southport stabbings. I believe this rationale justifies a legislative focus on the stirring up of extreme negative emotions, sentiments, or attitudes. The stirring up of mere dislike, mild suspicion, slight anxiety, or minor complaint would not be covered. This reform to the England and Wales

78. *See supra* note 57.

79. *See also* BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3, at ch. 3; Brown, *The "Who?" Question in the Hate Speech Debate: Part 2*, *supra* note 1, at 26-33; and Neller, *supra* note 45.

stirring up hatred offenses would reflect an important insight about hate speech: "that even though a good deal of hate speech is connected with emotions, feelings, or attitudes of hate or hatred, this is neither inevitable nor necessarily true of all instances of hate speech."[80] I use the phrase the "myth of hate" to describe the false assumption that hate speech is essentially about hate or hatred. In fact, hate speech can be motivated by, express, or stir up an almost unlimited range of extreme negative emotions, sentiments, or attitudes. These might include, but are not limited to, intense dislike, hatred, loathing, contempt, disdain, extreme aversion, extreme suspicion or mistrust, callous indifference, indignation, grievance, resentment, revenge, anger, rage, fear, alarm, disgust, revulsion, disapprobation, abhorrence, scorn, derision, serious envy, pity, superiority, profound disrespect, bigotry, prejudice, intolerance, rejection, or an exclusionary and/or discriminatory attitude.

Furthermore, I propose that the England and Wales stirring up hatred offenses be reformed to bring them closer into line with the International Covenant on Civil and Political Rights (ICCPR), the world's first dedicated international hate speech instrument.[81] Article 20(2) of the ICCPR calls on states parties to undertake the following legislative and judicial action: "Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law."[82] This is already reflected in the way hate speech laws are framed in some countries. Consider Art. 137d(1) of the Penal Code of the Netherlands:

> He who publicly, whether orally, in writing, or with imagery, incites hatred, discrimination or violent action against a person or belongings of people because of their race, their religion or their life philosophy, their gender, their heterosexual or homosexual orientation or their physical, psychological or mental disability, shall be punished by imprisonment of no more than a year or a monetary penalty [. . .].[83]

Along similar lines, I propose that the England and Wales stirring up hatred offenses are widened to also include the phrase "or incitement to discrimination or violence."

However, reacting to my proposal to widen the scope of the offenses to cover the stirring up of other sorts of extreme negative emotions, sentiments, or attitudes, and also incitement to discrimination or violence,

---

80. Brown, *What Is Hate Speech? Part 1*, *supra* note 4, at 440.
81. *See* BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3, at ch. 4.
82. Article 20(2) of the International Covenant on Civil and Political Rights (ICCPR).
83. Art. 137d(1) of the Penal Code of the Netherlands.

some may insist that this proposal simply will not do by itself. They will argue that the response misses an important point about viewpoint discrimination: that it can come in many different forms. With this in mind, let us now consider another form, exemplified in the following pair of examples.

Pairwise Comparison 2:

Not permitted: A person who uses threatening words or behavior, publicly, with intent to stir up hatred against Black people.

e.g. [words] "Black people are involved in a conspiracy to replace white people and so deserve our hatred; we should be prepared to defend white people with violence if necessary and the fact that they force us into this position is yet another reason to hate them."

e.g. [behavior] A person sees some people waiting to board a bus he is travelling on. He looks at everyone on the bus, and then walks to the entrance of the bus and allows a white person to get on but then blocks the way of a Black person, standing with his fists clenched, staring down at the person and then staring at the other passengers and shaking his head.

Permitted: A person who uses threatening words or behavior, publicly, with intent to stir up hatred against women.

e.g. [words] "Women are liars and sluts by nature, so guys you need to control your women. A wee slap here and there works just fine."

e.g. [behavior] Someone is giving the first-ever speech by a woman at an all-male club and a male member of the club walks on stage and physically ushers the woman off the stage mid-speech, to the cheers and applause of some of the men in the audience.

This second pairwise comparison arguably involves viewpoint discrimination because the law has the effect of restricting certain viewpoints on Black people while permitting viewpoints on women, for example. It gives the impression that the government is justifying the "racial hatred" element of the offense in terms of what it sees as "the wrongness of the underlying claim about race" and, therefore, that it is taking sides by implying a lack of wrongness of underlying claims about gender.[84] To repeat the words of the U.S. Supreme Court: "The government may not regulate use based on hostility—or favoritism—towards the underlying message expressed."[85] What is more, as well as the viewpoint of the speaker, there is also the viewpoint that the speaker

84. Ryu, *supra* note 69, at 249.
85. R.A.V. v. City of St. Paul, at 386.

is seeking to stir up or elicit in the minds of audience members, such as hatred of Black people or hatred of women. By banning the stirring up of hatred of Black people but permitting the stirring up of hatred of women, the government could appear to be criminalizing acts based on its disavowal of one viewpoint, to wit, hatred of Black people, and acceptance of another viewpoint, hatred of women.

There are two viable responses here. The first accepts that there is viewpoint discrimination at play but insists this can be justifiable on some valid basis. There may be such a thing as fair viewpoint discrimination, after all.[86] Consider the following valid basis articulated by the U.S. Supreme Court in *R.A.V.*: "When the basis for the content discrimination consists entirely of the very reason the entire class of speech at issue is proscribable, no significant danger of idea or viewpoint discrimination exists."[87] Suppose one of the main reasons that stirring up hatred against a group in society is proscribable has to do with the way this conduct indirectly heightens the risk of public disorder (e.g. race riots).[88] Let us also set aside for argument's sake the fact that, in relation to First Amendment jurisprudence at least, hate speech is a category of protected speech; the test for proscribable incitement is inciting "imminent lawless action"; and the test for proscribable threats is the existence of "true threats." Even if it is viewpoint discrimination to only ban the stirring up of hatred on grounds of race, religion, or sexual orientation—as is the case under the existing England and Wales legislation (but with a longer list operative in Scotland)—this could potentially be justifiable nonetheless, provided that these subcategories of stirring up hatred are especially serious or dangerous examples of the entire class of speech acts at issue. In other words, so long as the prohibition really is about the especially serious or dangerous subcategories of stirring up hatred, then only banning the stirring up of hatred on grounds of race, religion, or sexual orientation is not unfair viewpoint discrimination.

Of course, the discussion does not end there. Some might object that the proposed valid basis is a sham because it is unclear whether these particular subcategories of stirring up hatred really are especially serious or dangerous. They might argue that stirring up hatred on grounds of gender identity or disability is no less serious or dangerous. Therefore, if the government targets only the stirring up of hatred on grounds of race,

---

86. *See also* Bishop, *supra* note 70.

87. R.A.V. v. City of St. Paul, at 388.

88. *See also* BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3, at ch. 3; Brown, *The "Who?" Question in the Hate Speech Debate: Part 2*, *supra* note 1, at 26-33; Neller, *supra* note 45.

religion, or sexual orientation, then this is merely an excuse or sleight of hand, whereas the government's real aim is to suppress forms of speech it simply finds most disagreeable or inconvenient. This accusation is likely to emerge in situations where the government is not making known to the public both the body of evidence and standards of weighing evidence it is using to decide what are especially serious or dangerous subcategories of stirring up hatred. If the public is not informed whether heightened risk factors for public disorder are the quantity of stirring up of hatred, or the particular quality of stirring up of hatred or the specific social context—such as enmity between certain social groups—or all of the above, then the decision by the government to control only some subcategories of stirring up hatred will seem arbitrary. As a result, the public may harbor the suspicion that it is the government's own hostility towards some viewpoints, and tolerance of other viewpoints, that accounts for its regulatory approach.

Nevertheless, the danger or, more importantly, the public perception of the danger of the government's back-door suppression of those viewpoints it disfavors might recede if the government is more transparent with the public, and especially so if it undertakes a public consultation such that at least people might feel as though the government's decision is responsive to their will. Along these lines, one could argue that when the Coalition and Conservative governments, respectively, asked the Law Commission to conduct public consultations and reports on extending the stirring up hatred offenses both in 2013 (with the final report published in 2014)[89] and in 2018 (with the final report published in 2021),[90] this was partly to do with being transparent with the public about what body and standards of evidence they rely on when making this decision (good or bad), and partly to involve the public in the identification of relevant evidence and the formation of evidence-based opinion on the matter.[91] (A less charitable reading is that the government sought to kick the issues into the long grass or to effectively contract out the issue to a quango so as to avoid or at least disperse the blame for a controversial decision.)

---

89. LAW COMMISSION, HATE CRIME: SHOULD THE CURRENT OFFENCES BE EXTENDED?, LAW COM NO 348 (2014), https://lawcom.gov.uk/project/hate-crime-completed-report-2014/.

90. LAW COMMISSION, *supra* note 54.

91. *See* Alexander Brown, *People with Disabilities, Transgendered Identities, and Women Still Fair Game for Hate Speech Online*, HUFFINGTON POST U.K. (June 3, 2017), https://www.huffingtonpost.co.uk/dr-alexander-brown/people-with-disabilities-4_b_10273714.html; *and* BROWN & SINCLAIR, HATE SPEECH FRONTIERS, *supra* note 4, at 389.

A second viable response to pairwise comparison 2 accepts the force of the accusation of viewpoint discrimination and, furthermore, concedes that there is no valid basis for the differential treatment because the evidence does not support it, and so draws the only remaining conclusion, namely that the law should be amended to extend the list of protected characteristics appropriately. There are different ways of doing this. One is piecemeal: to extend the list of protected characteristics to include gender identity only. This change in the legislation would mean it would no longer be permitted to use threatening words or behavior with intent thereby to stir up hatred against women, for example.[92] But what about other protected characteristics and vulnerable groups? What about the characteristic of physical and mental capacity and disabled people as a group?[93] If the law does not permit stirring up hatred against women but does permit it for disabled people, then this remains viewpoint discrimination. Indeed, the list of protected characteristics needed to avoid viewpoint discrimination would seem to be very long. A law that covers race and some other recognized protected characteristics still does not cover lots of other things. In the words of Justice Scalia in *R.A.V.*, "Those who wish to use 'fighting words' in connection with other ideas—to express hostility, for example, on the basis of political affiliation, [or] union membership [. . .]—are not covered."[94] The list of protected characteristics could be extended to swallow up examples one by one. But this would require coming back to the law again and again to amend it, swallowing up parliamentary time. The specter of identity-based viewpoint discrimination would stalk the legislation.

Therefore, a bolder strategy that I recommend is to rewrite the law to make the list of protected characteristics formally open-ended. There is some precedent for this within international hate speech instruments as well as in some relevant domestic law. For example, the European Commission against Racism and Intolerance (ECRI) offers a definition of hate speech as part of its General Policy Recommendation No. 15 on combating hate speech, which provides a list of protected characteristics that ends with the open-ended clause "and other personal characteristics

---

92. *See* Brown, *The "Who?" Question in the Hate Speech Debate: Part 2*, *supra* note 1, at 26-33; Neller, *supra* note 1.

93. Alexander Brown, *New Evidence Shows Public Supports Banning Hate Speech Against People with Disabilities*, THE CONVERSATION (Mar. 1, 2017), https://theconversation.com/new-evidence-shows-public-supports-banning-hate-speech-against-people-with-disabilities-73807; and Brown, *The "Who?" Question in the Hate Speech Debate: Part 2*, *supra* note 1, at 26-33.

94. R.A.V. v. City of St. Paul, at 391.

or status."[95] In a similar vein, in 2022 the Committee of Ministers of the Council of Europe issued a Recommendation on combating hate speech that refers to "personal characteristics or status *such as* 'race', colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity and sexual orientation" (emphasis added)."[96] According to the Explanatory Memorandum to the Recommendation, "The list of grounds is purposefully open-ended."[97] At the domestic level, Finland has on the statute books a species of hate speech law that refers to stirring up hate on the basis of "race, colour, birth, national or ethnic origin, religion or belief, sexual orientation or disability *or on another comparable basis*" (emphasis added).[98] By embracing an open-ended list of protected characteristics, these instruments and laws bring hate speech laws much closer to the older category of seditious libel law, which is broadly related to stoking hostility and ill-will between any social classes or groups in a manner likely to produce a breach of the peace.

Once again, however, some may still object that these reforms are insufficient to reasonably mitigate against viewpoint discrimination. In particular, they might say there is a fatal asymmetry in pairwise comparison 2, namely that stirring up hatred against women is not the opposite of stirring up hatred against Black people. There is such a thing as standing up to, or even stirring up hatred against, racist bigots that is part of the tradition of public discourse in the U.K. and elsewhere, and, as written, the England and Wales stirring up hatred offenses permit that viewpoint, while not permitting the viewpoint of the racist bigot. Let us turn then to the third pair of examples.

Pairwise Comparison 3:

Not permitted: A person who uses threatening words or behavior, publicly, with intent to stir up hatred against Black people.

---

95. ECRI, GENERAL POLICY RECOMMENDATION NO. 15 ON COMBATING HATE SPEECH (Dec. 8, 2015), https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/recommendation-no.15.

96. RECOMMENDATION OF THE COMMITTEE OF MINISTERS OF THE COUNCIL OF EUROPE TO MEMBER STATES ON COMBATING HATE SPEECH, CM/REC. (2022) 16 (May 20, 2022), https://www.coe.int/en/web/combating-hate-speech/recommendation-on-combating-hate-speech.

97. EXPLANATORY MEMORANDUM TO THE RECOMMENDATION OF THE COMMITTEE OF MINISTERS ON COMBATING HATE SPEECH, para. 19 (2022), https://www.coe.int/en/web/combating-hate-speech/recommendation-on-combating-hate-speech.

98. Finnish Criminal Code, Ch. 11, §10, https://www.finlex.fi/api/media/statute-foreign-language-translation/511348/mainPdf/main.pdf?timestamp=1889-12-19T00%3A00%3A00.000Z.

e.g. [words] "Black people are involved in a conspiracy to replace white people and so deserve our hatred; we should be prepared to defend white people with violence if necessary and the fact that they force us into this position is yet another reason to hate them."

e.g. [behavior] A person sees some people waiting to board a bus he is travelling on. He looks at everyone on the bus, and then walks to the entrance of the bus and allows a white person to get on but then blocks the way of a Black person, standing with his fists clenched, staring down at the person and then staring at the other passengers and shaking his head.

Permitted: A person who uses threatening words or behavior, publicly, with intent to stir up hatred against racist bigots.

e.g. [words] "Death to racist scumbags!"

e.g. [behavior] A person stands in front of a crowd of anti-fascists assembled outside a pub in which a far-right organization is meeting, and the person paces up and down, slamming his fist into his own hand, pointing to the pub, shaking his head, and looking at the crowd.

Like the previous sets of examples, pairwise comparison 3 gives the impression that the government is deciding which viewpoints are acceptable for the public rather than allowing ordinary people to decide for themselves. It is not permitting the viewpoint of racists, but it is permitting the viewpoint of anti-racists.

One potential response to this pairwise comparison is once again to accept its force and amend the law accordingly. As already hinted at above, theoretically, the England and Wales stirring up hatred offenses could be extended to cover political beliefs and affiliations as a protected characteristic, and in theory being a racist bigot could be covered under the protected characteristic of political beliefs and affiliations. It is certainly not unimaginable that, in certain social contexts, a person stirring up hatred against a group of people on grounds of their political beliefs and affiliations could be especially serious or dangerous when the targets are an already oppressed group.[99] This change to the law would mean anti-racists are not permitted to stir up hatred against racist bigots, thus diffusing the viewpoint discrimination complaint.

Clearly there will be people who view the prospect of such an extension to the law with horror or consternation on free speech grounds, because it simply means yet more censorship. Interestingly, during a

99. *See* Brown, *The "Who?" Question in the Hate Speech Debate: Part 2*, *supra* note 1; BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3; and BROWN & SINCLAIR, HATE SPEECH FRONTIERS, *supra* note 4, at 109-12.

parliamentary debate on the Racial and Religious Hatred Bill in 2005, Tony Wright MP expressed a similar concern that the bill, as written, and without further amendments to protect freedom of expression, could end up criminalizing the stirring up of hatred against "religious bigots," given that in theory being a religious bigot could be covered under the protected characteristic of religion. He stated:

> I hate bigotry. I hate religious bigotry. All decent people should hate bigotry. I would like to incite people to hate bigotry, and I am worried about provisions that say that I cannot go round inciting people in that way. That incitement—which, as we have heard, involves loathing and intense dislike—is integral to our tradition.[100]

No doubt he would have leveled the same objection against a change in the law that made political beliefs and affiliations a protected characteristic and made it impermissible to stir up hatred against racial bigots. Extending the legislation in this direction might be considered by some a step too far, or even a *reductio ad absurdum* against hate speech laws in general. It is certainly likely to be politically unpopular among a section of society that includes anti-fascists and people who would, on principle, defend the free speech rights of anti-fascists. But what if there is clear evidence to suggest that this mode of stirring up hatred also carries a high risk of public disorder? Imagine a speaker shouting "Reform Party supporters are dirty racist bigots who should be met with force" to an angry mob of protestors affiliated with anti-fascist political parties or movements. If the government of the day refuses to extend the law in this way despite the existence of evidence of a type it has previously cited for other protected characteristics, then people will again reasonably suspect that this is because of the government's disdain for the viewpoint associated with stirring up hatred against Black people but approval for the viewpoint associated with stirring up hatred against racists and other types of bigots.

Nevertheless, a second response rejects the force of pairwise comparison 3 and argues that it is not evidence of viewpoint discrimination after all. It also resists the drive to extend the law to cover political beliefs and affiliations. There is no viewpoint discrimination here because, so the response runs, the existing England and Wales stirring up hatred offenses both permit anti-racists to stir up hatred against racists and permit racists to stir up hatred against anti-racists. Because the law allows both viewpoints tit for tat, it cannot be accused of taking sides. After all, under these existing offenses, there is nothing prohibiting a racist or some

100. HC Deb (June 21, 2005) cols. 729.

other type of bigot from shouting to a crowd of protestors: "It is time to liberate our great nation from racial egalitarians, by force if necessary." Nor do these offenses prohibit someone posting online the following sentence: "I hate so-called anti-fascists, and you should be honest about how much you hate them too, and when you are they had better watch out!"

However, there will be people who still think that pairwise comparison 3 misses the point and asymmetry remains in these examples. Stirring up hatred against racist bigots is not the opposite of stirring up hatred against Black people. So far, all the examples involve stirring up *negative* emotions, sentiments, or attitudes, and this, some will say, tells you everything you need to know about why the England and Wales stirring up hatred offenses involve viewpoint discrimination. Even if these offenses were redrafted so as to restrict stirring up hatred against *any group whatsoever*, the offenses would still restrict stirring up hatred whilst permitting stirring up *positive* emotions, sentiments, or attitudes. Putting this point another way, it might be tempting to say, "Stirring up hatred is a viewpoint," to echo Justice Alito's dictum from *Matal v. Tam*, "Giving offense is a viewpoint."[101] (In that case, Simon Tam, the lead singer of the rock group "The Slants", sought to register the term as a trademark in order to "reclaim" the term and dissolve its denigrating force as a slur for Asian persons. The Patent and Trademark Office (PTO) denied the application under the Lanham Act's disparagement clause prohibiting the registration of trademarks that may "disparage [. . .] or bring [. . .] into contemp[t] or disrepute" any "persons, living or dead."[102]) Reflecting this objection, we move to the final pair of examples.

Pairwise Comparison 4:

Not permitted: A person who uses threatening words or behavior, publicly, with intent to stir up hatred against Black people.

e.g. [words] "Black people are involved in a conspiracy to replace white people and so deserve our hatred; we should be prepared to defend white people with violence if necessary and the fact that they force us into this position is yet another reason to hate them."

e.g. [behavior] A person sees some people waiting to board a bus he is travelling on. He looks at everyone on the bus, and then walks to the entrance of the bus and allows a white person to get on but then blocks the

101. Matal v. Tam, 582 U.S. 218.
102. 15 U.S.C. §1052(a).

way of a Black person, standing with his fists clenched, staring down at the person and then staring at the other passengers and shaking his head.

Permitted: A person who uses threatening words or behavior, publicly, with intent to promote racial tolerance and love toward Black people.

e.g. [words] "This country is in the grip of intolerance towards Black people based on lies about replacement, and our willingness to use force may be the only way to bring people to their senses and come to tolerate or even love Black people as they deserve."

e.g. [behavior] A person sets up camp on the street in front of a television company and everyday sets on the ground a rose and a knife.

The thrust of pairwise comparison 4 is that if a law makes it permissible to promote tolerance and love but not intolerance and hate, then that law involves viewpoint discrimination. This is certainly the opinion of Weinstein:

> Unlike a ban on fighting words or profanity [. . .], hate speech bans are inherently viewpoint discriminatory. Britain's hate speech law, for instance, restricts only speech that intends to "stir up racial hatred" but not expression promoting racial tolerance. As a result, the discriminatory effect of hate speech laws persists even if the scope of the ban is confined to vituperation.[103]

In fact, Weinstein's general point about laws distinguishing between love and hate being viewpoint-discriminatory is one that has been made by the U.S. Supreme Court in several cases.

> The record confirms that any distress occasioned by Westboro's picketing turned on the content and viewpoint of the message conveyed, rather than any interference with the funeral itself. A group of parishioners standing at the very spot where Westboro stood, holding signs that said "God Bless America" and "God Loves You," would not have been subjected to liability. It was what Westboro said that exposed it to tort damages.[104]

> The First Amendment's viewpoint neutrality principle protects [. . .] the right to create and present arguments for particular positions in particular ways, as the speaker chooses. By mandating positivity, the [disparagement clause of the Lanham Act] might silence dissent and distort the marketplace of ideas.[105]

---

103. Weinstein, *Hate Speech Bans, Democracy, and Political Legitimacy*, *supra* note 3, at 545.

104. *Snyder*, 562 U.S. at 457.

105. *Matal*, 582 U.S. at 218, (Kennedy, J., concurring).

> So, the Lanham Act allows registration of marks when their messages accord with, but not when their messages defy, society's sense of decency or propriety. Put the pair of overlapping terms together and the statute, on its face, distinguishes between two opposed sets of ideas: those aligned with conventional moral standards and those hostile to them; those inducing societal nods of approval and those provoking offense and condemnation. The statute favors the former, and disfavors the latter. "Love rules"? "Always be good"? Registration follows. "Hate rules"? "Always be cruel"? Not according to the Lanham Act's "immoral or scandalous" bar. The facial viewpoint bias in the law results in viewpoint-discriminatory application.[106]

Once again, however, some credible responses are available to the charge that pairwise comparison 4 is evidence of viewpoint discrimination. One response goes back to basics by asking: What is a viewpoint? The gloss given by the U.S. Supreme Court is that opinions, perspectives, messages, and ideological beliefs can all be viewpoints. Scholars who believe that hate speech laws like the England and Wales stirring up hatred offenses involve viewpoint discrimination sometimes define the term "viewpoint" in a reasonably narrow way. As Eric Heinze writes, "I shall limit the concept of viewpoint to more-or-less general opinions."[107] But, arguably, stirring up hatred is not itself a viewpoint, it is conduct, and qua conduct the jurisprudential idea of viewpoint discrimination does not apply to it, not even hypothetically. In suggesting that stirring up hatred is not a viewpoint, I am deliberately parting company with Justice Alito's dictum, "Giving offense is a viewpoint." By restricting stirring up hatred, therefore, the government is not disfavoring one side of a debate. Rather, it is disfavoring a specific type of conduct, which it does anytime it criminalizes activity. In other words, it cannot be viewpoint discrimination to restrict a specific class of conduct because the concept of viewpoint discrimination is only applicable to restrictions of speech.

In fairness to Heinze, when writing about whether hate speech laws involve viewpoint discrimination, he attends to the distinction between the regulation of speech and conduct. For example:

> A working distinction between speech and conduct indicates some of the limits on speech that remain compatible with viewpoint absolutism. For example, the U.S. Supreme Court recognizes legitimate, non-viewpoint-based regulations, i.e., incidental restrictions on speech, arising under

106. Iancu v. Brunetti, 588 U.S. 388, 390 (2019).

107. Heinze, *Viewpoint Absolutism and Hate Speech*, *supra* note 3, at 548.

> 'time, manner and place' restrictions, such as banning megaphones in residential areas late at night, or requiring permits for demonstrations in crowded city areas. Such restrictions are legitimate insofar as they regulate only the conduct component, and not the viewpoint component, of the speech, and do not become sham pretexts for viewpoint regulation. Line drawing problems can certainly arise—Was a ban imposed because of the megaphone or the message?—but the relevant evidentiary problems concern any form of de facto discrimination, and any form of controversial speech. Hate speech introduces no additional difficulties.[108]

However, to say that "Hate speech introduces no additional difficulties" belies the fact that Heinze himself assumes that all hate speech laws regulate not the conduct component, but instead the viewpoint component of what is proscribed. I reject that simple assumption.

In fact, there has been a tendency among many free speech scholars to make a generalization about hate speech laws, specifically that, unlike time, place, and manner restrictions, they are primarily, or for all intents and purposes, speech-restricting as opposed to conduct-restricting. According to Weinstein, for example, if viewpoint discrimination also means restricting speech based on the specific worldview it expresses, then the doctrine is straightforwardly applicable to incitement to hatred laws.[109] Suppose these laws restrict the use of racist speech to stir up hatred against Black people and that an example of racist speech would be the following: "America is being swamped by coloreds who do not believe in democracy and harbor a hatred for white people." Weinstein writes, "Racist speech such as this, although expressing an ugly, twisted view of the world, does nonetheless express a worldview."[110] Yet this overlooks the point that what incitement to hatred laws directly prohibit is the act of inciting hatred and only indirectly prohibit the speech, such as racist speech, used to facilitate that class of conduct. Of course, laws banning Holocaust denial might be a different kettle of fish. These laws also directly prohibit the act of denialism and indirectly prohibit the words used to perform that act. Yet, laws banning Holocaust denial certainly appear to be primarily speech-restricting as opposed to conduct-restricting. However, given that hate speech laws are a heterogeneous collection of laws, to assume that all such laws are primarily speech-

---

108. *Id.* at 571.

109. Weinstein, *Hate Speech, Viewpoint Neutrality, and the American Concept of Democracy*, *supra* note 3, at 152.

110. *Id.*

restricting as opposed to conduct-restricting would be an unsustainable over-generalization.[111]

So, the pressing question is this: Are the England and Wales stirring up hatred offenses primarily speech-restricting or conduct-restricting? To answer this question, we need to lean into the speech/illegal conduct distinction, which is an even older doctrine in First Amendment jurisprudence than the viewpoint discrimination doctrine. The core idea behind this distinction can be captured in the famous dictum from *Schenck v. United States*,[112] "The most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic."[113] The more general thought here is about a class of illegal conduct that is frequently, *but does not have to be*, performed using speech. This is the sort of thing Justice Scalia had in mind when he wrote the following in *R.A.V.*:

> since words can in some circumstances violate laws directed not against speech but against conduct (a law against treason, for example, is violated by telling the enemy the Nation's defense secrets), a particular content-based subcategory of a proscribable class of speech can be swept up incidentally within the reach of a statute directed at conduct rather than speech.[114]

Arnold H. Loewry offers a longer list of similar examples:

> The robber may or may not say, "your money or your life" when he points his gun at the victim. The murderer may or may not suggest or command that his cohort fire the fatal bullet. The prankster seeking to cause panic may either start a fire in a crowded theater or simply shout "fire." None of these criminals are protected simply because they used words to commit their crimes.[115]

Cass Sunstein even proffers cases where "words actually amount to a way of performing independently illegal acts" of racial discrimination.

> If someone writes a letter saying, "You're fired, because I won't let blacks work here," we can properly categorize the letter as a form of action. The letter amounts to a commission of an illegal act, that of racially discriminatory discharge. If government can punish that act, surely it can punish the speech that is that act. The letter is simply evidence of what is

111. *See* BROWN, HATE SPEECH LAW, *supra* note 2, at ch. 2.

112. 249 U.S. 47 (1919).

113. *Id.* at 52.

114. R.A.V. v. City of St. Paul, at 389.

115. Arnold H. Loewry, *Distinguishing Speech from Conduct*, 45 MERCER L. REV. 621, 622 (1993).

> unlawful, a discharge based on discrimination. Use of the letter to prove discriminatory motive is hardly unconstitutional even if the letter is speech.[116]

According to this model, certain actions are treated as illegal conduct despite being performed with words, meaning the regulation targets the conduct and restricts the words as an incidental consequence. Importantly, this model can be, and has been, applied to many kinds of hate speech laws. For example, it is the model that, *pace* Justice Scalia, Justice White applied to the cross-burning ordinance in *R.A.V.*:

> The majority's observation that fighting words are "quite expressive indeed," [. . .] is no answer. Fighting words are not a means of exchanging views, rallying supporters, or registering a protest; they are directed against individuals to provoke violence or to inflict injury. [. . .] Therefore, a ban on all fighting words or on a subset of the fighting words category would restrict only the social evil of hate speech, without creating the danger of driving viewpoints from the marketplace.[117]

To give another example, Kenneth L. Marcus has argued that the idea of First Amendment protection is not "salient" to speech, which constitutes discriminatory harassment in the workplace or on a university campus because what is really at stake is not the speech itself but the illegal conduct the speaker uses the speech to perform. As he put it, "antidiscrimination law pulls in harassing campus speech only as an incidental constituent of behavior addressed under a well-established regulatory scheme."[118]

To offer a third example, according to Ishani Maitra and Mary Kate McGowan, when racist hate speech "marks people of color as socially subordinate (to whites), and legitimates discriminatory behavior towards them," it "enact[s] changes in [social] obligations towards people of color," including "obligations to perform illegal acts (i.e., acts of racial discrimination)."[119] Therefore, in their view, this sort of racist hate speech qua conduct (i.e. the action of enacting social obligations) "should not be covered by the First Amendment."[120]

An important assumption underpinning all these examples is the principle that, if there exists a compelling state interest in treating certain

---

116. SUNSTEIN, *supra* note 3, at 125-26.

117. *R.A.V.* 112 U.S. at 401 (White J., concurring).

118. Kenneth L. Marcus, *Higher Education, Harassment, and First Amendment Opportunism*, 16 WM & MARY BILL RTS J. 1025, 1057 (2008).

119. Maitra & McGowan, *supra* note 3, at 369.

120. *Id.*

acts as illegal conduct, then this interest carries over to the speech that is frequently, but need not be, used to perform the illegal conduct, even where regulating the conduct would regulate the speech as an incidental consequence. I believe this principle also applies *mutatis mutandis* to the sort of illegal conduct that is the main subject of this Article, stirring up hatred. In particular, the England and Wales stirring up hatred offenses pertain to illegal conduct that is frequently, but need not be, performed using speech. Recall that the offenses apply, as written, to words *or behavior*.

Now, some might be immediately skeptical towards the claim that stirring up racial hatred, for example, is frequently, but need not be, performed using speech. They might think that, in practice, stirring up racial hatred *must* involve speech akin to how the crime of perjury requires speech. However, the England and Wales stirring up hatred offenses contain the phrase "words or behavior" for good reason. Relevant sorts of behavior might include: blackface, marching, standing outside premises day after day, loitering late at night, encroaching on people's personal space, faking a punch, making a gun or shooting gesture, burning a cross, deliberately bumping into people, pointing a finger or poking people with your finger, refusing to serve people in a shop, blocking people's path, physically ushering people off stages, locking people out of buildings, removing people's property and putting it onto the street, deliberately brandishing items that might be considered dangerous or mistaken for weapons, or repeatedly revving a car engine outside someone's home.

I have suggested that the doctrine that says governments should not engage in viewpoint discrimination disapplies to laws directed not against speech as such, but against conduct, even when an incidental consequence of making certain conduct illegal is to restrict certain speech. Nevertheless, there is another counterargument waiting in the wings. It starts with the premise that the viewpoint discrimination doctrine remains applicable even to laws that restrict conduct if the restriction of *either speech or expressive behavior* is more than an incidental consequence of the law. This occurs when the conduct cannot but be performed without either speech or expressive behavior. For the present purposes, the term expressive behavior can refer to behavior designed to convey a message that is likely to be understood by receivers, and which is thereby on a par with speech in terms of its expressive value or quality. Arguably, this is the position of the U.S. Supreme Court as demonstrated by its rulings in

cases involving the burning of the American flag, for example.[121] In the words of Justice Brennan, "if Texas means to argue that its interest does not prefer *any* viewpoint over another, it is mistaken; surely one's attitude toward the flag and its referents is a viewpoint."[122] In cases of stirring up hatred involving behavior, the behavior might be expressive insofar as it is designed and understood to send the message that a certain group is despicable and deserve to be hated or some other similar message (or so says the current counter-argument). Furthermore, some might argue that the crime of stirring up hatred, like the crimes of bribery or incitement, requires or necessarily involves either speech or expressive behavior. So, the idea of viewpoint discrimination is applicable. By analogy, some scholars have argued that any law banning drag performances would violate the First Amendment because drag performances are expressive behavior that express viewpoints and so are protected by the First Amendment.[123]

If this counterargument is correct, then we find ourselves circling back to the original objection that it is viewpoint discrimination to ban stirring up hatred whilst permitting promoting love and tolerance. There are two other ways of articulating this objection worth mentioning at this stage, as they reveal the full extent of the worry about the England and Wales stirring up hatred offenses. First, pairwise comparison 4 might create the impression that the government's underlying rationale for the law is in fact the specific emotion or sentiment being stirred up—hatred as opposed to love—and this smacks of blatant viewpoint discrimination. As Michael Korengold puts it,

> Prohibiting incitement to racial hatred would punish a person for causing another to hate, a result which is dangerously close to prohibiting the thought or feeling of hatred itself. Criminalization of a specific thought or feeling is precisely the type of viewpoint regulation that compromises the right to free expression.[124]

Second, pairwise comparison 4 also might imply that the government's underpinning rationale for the law is the specific ideological motivation of the perpetrator—such as a belief in racial superiority, an attitude of racial prejudice, or a sentiment of racial

---

121. *See* Spence v. Washington, 418 U.S. 405 (1974); Texas v. Johnson, 491 U.S. 397 (1989).

122. *Id.* at 414.

123. Mark Satta, *Shantay Drag Stays: Anti-Drag Laws Violate the First Amendment*, 25 GEO. J. GENDER & L. 95 (2023).

124. Korengold, *supra* note 69, at 727.

intolerance as contrasted with a belief in racial equality, an attitude of racial impartiality, or a sentiment of racial tolerance—and this once again looks like viewpoint discrimination. To repeat the words of Justice Kennedy in *Rosenberger*, "The government must abstain from regulating speech when the specific motivating ideology or the opinion or perspective of the speaker is the rationale for the restriction."[125]

Of course, being sensitive to these worries, the government could speak clearly and plainly to the public about what the stirring up hatred offenses do and do not criminalize and, relatedly, what the rationale behind these laws is. First, the government could emphasize that the offenses do *not* criminalize the thought or feeling of hatred itself, but rather the stirring up of hatred. The fact that prohibiting the stirring up of hatred comes dangerously close to prohibiting the thought or feeling of hatred itself does not mean they are the same thing, because they are not.[126] Second, the government might clarify that the offenses do *not* criminalize motivational bases behind the stirring up of hatred, but rather they criminalize the use of certain types of words or behavior to stir up hatred. This is distinct from hate crimes, where motivational bases can be elements of the relevant offenses.[127] Most importantly, the government could make a series of statements in parliament and through the media concerning the real rationale for the law, whether it is the threat the offending behavior poses to public order or something else.[128] It is unclear whether making these public statements will be enough to persuade the whole of society that no viewpoint discrimination is involved, but they could be a useful step in that direction.

Nevertheless, what would it take to make the risk of viewpoint discrimination recede entirely? Or, as Justice Scalia puts it in *R.A.V.*, how might a government restrict speech in a manner "such that there is no realistic possibility that official suppression of ideas is afoot."[129] A radical approach would be to make it illegal to use threatening words or behavior, publicly with intent and being likely thereby to stir up any extreme emotions, sentiments, or attitudes, *whether they be negative or positive*, from an open-ended list, provided that it is against other persons on the grounds of an open-ended list of protected characteristics. The

---

125. Rosenberger v. Rector and Visitors of University of Virginia, 515 U.S. 819, 829 (1995).

126. *See also* BROWN, HATE SPEECH LAW, *supra* note 2, at 26-27; and Alexander Brown, *What Is Hate Speech? Part 2: Family Resemblances*, 36 L. & PHIL. 561, 606-07 (2017).

127. *See also* BROWN & SINCLAIR, HATE SPEECH FRONTIERS, *supra* note 4, at ch. 6.

128. *See* BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3, at ch. 3.

129. R.A.V. v. City of St. Paul, at 390.

government could also publish an explanatory note stating that, in theory, the offense could cover cases of stirring up love, provided all the other elements are also present, even if, in all probability, there will be very few instances of this.

However, I reject this radical approach on several grounds. First, stirring up extreme negative or extreme positive attitudes towards other persons on the grounds of protected characteristics are not equal or proportionate activities. For example, stirring up extreme positivity might be significantly less likely to lead to public disorder. These are not the sorts of emotions, sentiments, or attitudes most often associated with violent protests, mass brawls, rioting, and assaults. In addition, suppose another rationale for laws banning the stirring up of hatred is to address the cumulative contribution this activity makes to a general and widespread attitude of hatred, contempt, prejudice, resentment, mistrust, fear, and anger towards minorities, where this "climate of hatred" involves an increased likelihood of acts of discrimination, damage to property, violence and other hate crimes against those minorities.[130] It is unclear that this legislative rationale warrants extending the offenses to cover the promotion of extreme positivity. Typically, extreme positivity does not contribute to a social climate characterized by a heightened risk of discrimination and violence, which, presumably, is one reason why society feels more comfortable about people trying to promote positivity.

Second, there might be a negativity bias in how people respond to the stirring up of extreme emotions, sentiments, and attitudes. In particular, it is possible that exposure to the stirring up of extreme negativity has a more significant impact on audiences' emotional and/or cognitive states of mind than exposure to the stirring up of extreme positivity. If the legislative purpose is ultimately to prevent or disincentivize the stirring up of extreme emotions, sentiments, and attitudes, then it might be legitimate to focus legislative firepower on the most potent or effectual forms of stirring up.

These initial points exemplify one of the generic, valid bases for viewpoint discrimination identified in *R.A.V.* In the present instance, the logic is that, when the basis for treating the stirring up of extreme negativity differently than the stirring up of extreme positivity consists entirely of the very reasons why the entire class of speech at issue is proscribable, then there is no significant danger of (unjustified) viewpoint discrimination.

---

130. *See* Brown, *The Racial and Religious Hatred Act 2006*, *supra* note 1, at 13-14; BROWN, HATE SPEECH LAW, *supra* note 2, at 66-71, 75; and BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3, at 318-19.

Third, there is the practical point that in order to justify the valuable legislative time and political capital needed to pass amendments to existing legislation, a case needs to be made that the phenomenon that the amendments are intended to address are real and prevalent.[131] Expending precious legislative resources on any given law (or amendment to existing law) comes at an opportunity cost in terms of other laws (or amendments) that could have been brought forward, so the cost needs to be worth it. But if in fact there are likely to be scarcely any cases involving a person using threatening words or behavior, publicly, with intent thereby to stir up love, and meeting all the other elements of the offense, then the amendment would seem inefficient.

Fourth, there might be a higher prevalence of cases of faux stirring up connected with the stirring up of positivity than with the stirring up of negativity. For example, there could be cases of sarcastic, insincere, or microaggressive uses of language in which the speaker appears to be stirring up positivity but is actually seeking to stir up negativity. Consider the following content a bigoted person might post on social media: "These black men are so strong, and big downstairs, that our white women all love them, and woe betide any women who doesn't fall in love with them." If some of the very few cases that appear to involve the stirring up of positivity in fact typically turn out to be faux cases, this would render amendments of legislation designed to capture such cases even less worthwhile.

Fifth, if the England and Wales stirring up hatred offenses were given a radical facelift such that they also covered the stirring up of love, then this would fundamentally change their character. The current offenses are part of a much larger family of laws that includes not only all incitement to hatred laws but also laws banning inciting hatred, discrimination, or violence; laws prohibiting inciting hatred in a way likely to cause a breach of the peace; group defamation laws; and even seditious libel laws. All these laws reflect, either directly or indirectly, the fundamental insight that certain conduct can contribute to contempt, enmity, hostility, and violence between different sections of society.

IV. A MODEL STIRRING UP OFFENSE

I believe that having a proper focus on legitimate concerns about stirring up hatred, as well as paying attention to all arguments relating to the viewpoint discrimination and overbreadth doctrines, leads to a

131. *See* Brown, *The "Who?" Question in the Hate Speech Debate: Part 1*, *supra* note 1, at 293-301.

different and better legal response than the current England and Wales stirring up hatred offenses. Drawing on all of these insights and arguments, I propose that the current offenses be substantially redrafted and brought under the umbrella of the following model stirring up offense:

> *It is an offense for a person to use threatening words or behavior, or to display, publish, or post threatening words or behavior, publicly (whether offline or online), with intent and with regard to all the circumstances being likely thereby to stir up extreme negative emotions, sentiments, or attitudes (including but not limited to hatred, contempt, scorn, prejudice, mistrust, grievance, resentment, revenge, anger, rage, or fear), or to incite discrimination or violence against another person or group of people on grounds of their possession or perceived possession of protected characteristics (including but not limited to race, religion, sexual orientation, gender identity, or ability).*

Of course, at this stage several other free speech worries—not to mention First Amendment doctrines—coming flooding back into the picture. Principal among them is that the model stirring up offense would probably still face an overbreadth challenge if, hypothetically speaking, it were brought before the U.S. Supreme Court. It might also face a similar sort of challenge if it were brought before the judges of the U.K. Supreme Court, and ultimately the ECtHR, especially if the judges are influenced in their thinking by the overbreadth doctrine. Potentially, the model offense could sweep up a range of speech that ought to be protected. For example, it could sweep up political speech, scientific speech, artistic speech, or even comedic speech, if this sort of speech happened to be used to facilitate the relevant kind of stirring up. It is worth noting that U.S. courts have previously struck down hate speech laws, as well as laws that have the effect of restricting hate speech, for being unconstitutionally overbroad.[132]

However, I believe the overbreadth challenge can be met. One general point to make here is that, as articulated (or rearticulated) by the U.S. Supreme Court in recent years, the overbreadth doctrine speaks to laws that, whilst aimed at restricting unprotected speech or speech that facilitates illegal conduct, are realistically going to sweep up a

---

132. Collin v. Smith II, 578 F.2d 1197 (7th Cir. 1978); Doe v. University of Michigan, 721 F. Supp. 852 (E. D. Mich.1989); Virginia v. Black, 538 U.S. 343 (2003); State v. Turner, 864 N.W.2d 204 (Minn. Ct. App. 2015); Myers v. Fulbright, 367 F. Supp. 3d 1171 (D. Mont. 2019). For discussion of the overbreadth findings in some of these cases, see BROWN, HATE SPEECH LAW, *supra* note 2, at 264-66.

substantially disproportionate amount of protected speech. As such, it matters how much political speech, scientific speech, artistic speech, or comedic speech the model offense is realistically going to sweep up compared to the amount of speech that is legitimately in its crosshairs. It might be that, in reality, the amount of the former is not substantially disproportionate to the amount of the latter. (Once again, I set aside for the sake of making this particular argument the fact that the U.S. Supreme Court has identified hate speech as itself a category of protected speech.[133])

Furthermore, I highlight three aspects of the model offense that I believe would mitigate the tendency towards overbreadth, even if it would not eliminate the tendency completely, by making the offense narrower than it might otherwise be. First, the model offense reverts back to the wording of Section 6(1) of the Race Relations Act 1965 for all subcategories of stirring up hatred (race, religion, and sexual orientation), and specifically includes a conjunction of the elements of "intent to stir up hatred" and "likely to stir up hatred."[134] This would place a higher bar on prosecution and significantly reduces the risk of sweeping up speech that ought to be protected. It deserves mention here that, as it stands, within the jurisdiction of England and Wales, the Public Order Act 1986 includes these elements subjunctively ('intends thereby to stir up racial hatred, *or* [. . .] having regard to all the circumstances racial hatred is likely to be stirred up thereby', emphasis added), in the case of stirring up racial hatred (Part 3 of the Act), and only includes the intent element in the case of stirring up religious or sexual orientation hatred (Part 3A of the Act). By contrast, within the jurisdiction of Scotland, the Hate Crime and Public Order (Scotland) Act 2021 includes similar elements subjunctively, in the case of all subcategories of stirring up hatred (Part 3 of the Scottish Act):

> . . . in doing so, the person intends to stir up hatred against a group of persons based on the group being defined by reference to race, colour, nationality (including citizenship), or ethnic or national origins, *or* [. . .] a reasonable person would consider the behaviour or the communication of the material to be likely to result in hatred being stirred up against such a group (emphasis added).

It is also important to highlight that, under U.S. First Amendment doctrine, courts have treated several categories of speech as less protected or unprotected partly based on the fact that they include speaker intentions

133. *See supra* note 24.

134. § 6(1) of the Race Relations Act 1965.

as elements of the speech in a way that significantly narrows the scope of the speech, thereby reducing the risk that the relevant restrictions would be overbroad and sweep up—and chill—a substantially disproportionate amount of protected speech.[135]

Second, the model offense only pertains to the stirring up of "extreme negative" emotions, sentiments, or attitudes, as well as the incitement of discrimination or violence, against persons or groups of people on the grounds of protected characteristics. This qualification also narrows the field of cases that could be realistically subject to prosecution and thereby minimizes the risk of overbreadth.

Third, a person is guilty of the model offense only if they use "threatening words or behavior" to perform the relevant stirring up. Here, again, by narrowing the offense, there is less chance it will swallow up a substantially disproportionate amount of protected speech.

That said, it might be objected at this stage that there is lacking an independent rationale for limiting the offense to "threatening words or behavior" other than simply to reduce the chances of overbreadth. Some might even regard this limitation as ad hoc. After all, if the principal aim were to restrict words or behavior that are likely to stir up hatred or other forms of extreme negativity in the relevant sense, then there might be good reasons not to draw the offenses so narrowly. For one thing, why not redraft the existing offenses of stirring up hatred on the grounds of religion or sexual orientation (set out in Part 3A of the Public Order Act 1986) to cover words that are not only "threatening" but also abusive or insulting, slurring, stereotyping, denigratory, disparaging, defamatory, demeaning, mocking, derisory, humiliating, or *any other* categories of words which are likely to stir up hatred?[136] Furthermore, to reflect the aim of mitigating against viewpoint discrimination, and the fact that my model offense pertains to the stirring up of any extreme negative emotions, sentiments, or attitudes, then why not also cover words that are hyperbolic, propaganda, distorting, disinformation, false narratives, wildly emotional, hysterical, or *any other* categories of words which are likely to stir up extreme negativity in the relevant sense?

Of course, a further ground for limiting the model offense to threatening words or behavior could be the fear of violence or reduced sense of personal security, as in, a lack of confidence in the safety of one's body and possessions, suffered by persons or groups against whom

---

135. *See* Leslie Kendrick, *Speech, Intent, and the Chilling Effect*, 54 WM & MARY L. REV. 1633 (2013).

136. *See also* Bindman, *Outlawing Hate Speech*, *supra* note 1, at 18; and BROWN & SINCLAIR, HATE SPEECH FRONTIERS, *supra* note 4, at 387.

extreme negative emotions, sentiments, or attitudes are being stirred up.[137] Indeed, under the bespoke public order law of Northern Ireland, the stirring up hatred offenses are couched in terms of a person using "threatening, abusive or insulting words or behaviour" with intent thereby to "stir up hatred or arouse fear" or "having regard to all the circumstances hatred is likely to be stirred up or fear is likely to be aroused thereby."[138] However, if this is a key rationale for the offenses, then there could be grounds for reforming them differently again. The existing stirring up hatred offenses could be supplemented with additional elements such as the following: that the speaker also had intent to cause people to feel threatened or to be fearful; that the words or behavior were also likely to cause other persons to feel threatened or to be fearful; or even that the words or behavior also thereby caused other persons to feel threatened or to be fearful. However, these reforms would need to be justified in the face of the counterargument that they would render the stirring up hatred offenses redundant.[139] After all, Section 2 of the Public Order Act 1986 already makes it an offense for a group of three or more persons to threaten unlawful violence in a manner that would cause a reasonable person present at the scene to fear for their personal safety.[140] Additionally, Section 4A of the same Act makes it an offense for a person to use threatening, abusive or insulting words or behavior with intent to cause another person harassment, alarm, or distress and thereby causing that other person harassment, alarm, or distress.[141] Moreover, Section 181 of the Online Safety Act 2003 makes it an offense to send a message by electronic means that conveys a threat of death or serious harm, and intending a person encountering the message to fear that the threat would be carried out, or being reckless as to whether an individual encountering the message would fear that the threat would be carried out.[142] Each of these offenses can be treated, at the very least, as hate crimes at the sentencing stage if found to be "aggravated by hostility" based on one of five protected characteristics under Section 66 of the Sentencing Act 2020.

Notwithstanding these points, there is one remaining reason for limiting the existing England and Wales stirring up hatred offenses and my model stirring up offense to only threatening words or behavior that I

137. *See* BROWN, HATE SPEECH LAW, *supra* note 2, at 71-75.
138. Part 3 of the Public Order (Northern Ireland) Order 1987.
139. *See also* BROWN & SINCLAIR, HATE SPEECH FRONTIERS, *supra* note 4, at 388.
140. § 2 of the Public Order Act 1986.
141. *Id.* at § 4A.
142. § 181 of the Online Safety Act 2003.

believe deserves greater attention than it has received among parliamentarians, namely the value of substantive autonomy. A person using threatening words or behavior to perform an act of stirring up extreme negative emotions, sentiments, or attitudes might diminish or violate the audience's autonomy given facts about the speaker, the audience, the speaker-audience relationship, the wider social context, and/or the specific speech situation.[143] I believe the value of substantive autonomy can provide a bespoke and robust supplemental rationale for limiting the relevant offenses to only "threatening words or behavior," especially offenses whose gravamen is stirring up any extreme negative emotions, sentiments, or attitudes against persons or groups of people based on protected characteristics.

To understand why, I believe there are lessons to be learned here of John Stuart Mill's well-known corn-dealers example:

> An opinion that corn-dealers are starvers of the poor, or that private property is robbery, ought to be unmolested when simply circulated through the press, but may justly incur punishment when delivered orally to an excited mob assembled before the house of a corn-dealer, or when handed about among the same mob in the form of a placard.[144]

If the sole worry were a breach of the peace and/or the safety of corn-dealers, then it is hard to see why these negative opinions would be dangerous when delivered orally to an excited mob but *not* dangerous when simply circulated through the press. However, an additional concern could be about the threat to the autonomy of the excited mob, who, due to crowd psychology, are more vulnerable to being emotionally manipulated by the speaker than individuals at home calmly reading newspapers.[145]

I suggest that, if not identical, then certainly a similar line of reasoning could be applied to both the England and Wales stirring up hatred offenses and my model stirring up offense. In particular, it could be that a person using threatening words or behavior renders receivers less capable of rationally assessing the ideas, opinions, or messages used to facilitate the stirring up (e.g. the message that a certain group are despicable and deserve to be hated or some other similar message) and/or less capable of rationally assessing the persuasive element of the stirring up (e.g. that the speaker is seeking to persuade or incite the audience into

---

143. *See also* Brown, *The Racial and Religious Hatred Act 2006*, *supra* note 1, at 10-12.

144. JOHN STUART MILL, *On Liberty*, *in* UTILITARIANISM, ON LIBERTY, CONSIDERATIONS ON REPRESENTATIVE GOVERNMENT 123 (Everyman ed, 1972).

145. *See also* Brown, *The Racial and Religious Hatred Act 2006*, *supra* note 1, at 11.

hatred or similar). To see how, it is first necessary to address an ambiguity in the wording of the England and Wales stirring up hatred offenses, specifically an ambiguity in the relevant subject matter of the statutory phrase "threatening words or behaviour." Must the threatening words or behavior encode or imply a threat to the audience or to the targets against whom hatred is being stirred up or either? Consider the following examples.

> e.g. [words] "You'd better come to hate Muslims as much as I do or else you'll have me to answer to."
>
> e.g. [words] "You'd better come to hate Muslims or else they will overrun this country and change it forever."[146]

Arguably, in these examples, the encoded or implied threat is towards the audience. If so, then it could potentially elicit annoyance, alarm, or even fear in the receiver. If fear, then this emotion could circumvent or curtail rational deliberation. It could stop people from thinking things through properly; it could drive "hot" and impulsive thoughts and decisions. These thoughts or decisions could be more likely to involve cognitive biases, flawed heuristics, or fallacies, such as over-relying on preconceptions, assumptions, or stereotypes about people (about the speaker or the targets of speech or both), because they are driven by fear.[147]

Alternatively, the threat could be towards the target, as in, the group against whom hatred is stirred up. Consider some further examples.

> e.g. [words] "You think you can trust Muslims, think again, they are vile, backward, and dangerous people who deserve only our hatred, and when the good and reasonable people of this country are finally united in their hatred of this radical element in their midst, then Muslims had better watch out!"
>
> e.g. [words] "Surely you can see that these Muslims don't belong here. Be honest, don't you think that they are like rats or cockroaches. And what would you do if you found pests or vermin in your house? That's right, you would call in pest control to exterminate them. I want us to do the same for Muslims."[148]

---

146. *See* BROWN & SINCLAIR, HATE SPEECH FRONTIERS, *supra* note 4, at 384-86.

147. Ruchi Rathor, *The Psychology of Fear: Understanding Its Impact on Decision-Making*, MEDIUM (Aug. 2, 2023), https://medium.com/@ruchirathor_23436/the-psychology-of-fear-understanding-its-impact-on-decision-making-f40788f40ab4.

148. *Id.*

By using this sort of threatening language, the speaker might also produce emotions such as annoyance, alarm, or fear in the targets. The fear could chip away at people's sense of personal security. Of course, fear of violence and a reduced sense of personal security are common effects of hate speech that can themselves render hate speech proscribable (creating a "climate of fear").[149] But in the context of the current discussion, I am suggesting that threats can be used not merely to stir up of hatred, but also to evince extreme negative emotions or sentiments in the targets of the stirring up, including fear, and fear can diminish the targets' substantive autonomy such as by circumventing or curtailing rational deliberation. So, for example, if people are denigrated and threatened, and they know these words are used to stir up hatred against them in the hearts of third parties, all because of the color of their skin, then it could produce a cocktail of resentment, paranoia, self-doubt, and fear in the targeted people, and such emotions could drive thoughts and decisions that are panicked and ill-conceived. These observations dovetail with existing analyses found in the literature about how other common forms of hate speech, such as slurs, epithets, group defamation, negative stereotypes, dehumanizing comparisons, and provocations, can also pose a threat to substantive autonomy, even if the mechanism is not fear but other negative emotions or sentiments like self-loathing, disillusionment, or anger.[150]

All of that being said, if the current England and Wales stirring up hatred offenses were substantially redrafted or repealed and replaced with my model stirring up offense, which covers the stirring up of any extreme negative emotions, sentiments, or attitudes, then potentially other categories of words could end up in the legislative crosshairs, if one of the rationales is to protect the value of substantive autonomy. Why not also cover propaganda, indoctrinating discourse, false narratives, subliminal messages, radicalizing language, or *any other* categories of words that are likely to stir up extreme negativity in a manner that could threaten receivers' substantive autonomy? I concede the argumentative force of this rhetorical question but reject its implied conclusion. In doing so, I once again fall back on the requirement of preventing the offenses from becoming overbroad and sweeping up a substantially disproportionate amount of speech that ought to be protected. For this reason, I recommend limiting my model stirring up offense to only "threatening words or behavior." Nevertheless, I also reject the criticism that this limitation is ad

149. BROWN, HATE SPEECH LAW, *supra* note 2, at 71-75.

150. *See* Brison, *supra* note 33, at 328; Brink, *supra* note 33, at 138-140; Moles, *supra* note 33; and BROWN, HATE SPEECH LAW, *supra* note 2, at 62-66.

hoc, where "ad hoc" has a negative connotation. On the contrary, the limitation is justified by principle, the principle being the overbreadth doctrine. Or to be more exact, the limitation reflects a reasonable balance between the viewpoint discrimination and overbreadth doctrines, as opposed to sacrificing a principle merely for the sake of expediency.

Nevertheless, this does not mean a society may do nothing to combat forms of hate speech that do not involve threatening words or behavior. On the contrary, reflecting the fact that there are two concepts of hate speech, the ordinary and legal concepts (see Part I), there is more than one way to combat hate speech, depending on which concept is in play. A person using slurs, negative stereotypes, group defamatory, disparaging, or dehumanizing words to stir up hatred against people on the grounds of protected characteristics, for example, could be deemed to be engaging in hate speech in the ordinary sense of the term (non-legalistic). Social media platforms, could classify this as disallowed hate speech, for example.[151] Social media sanctions such as content removal or even account suspension could be another way private companies do their bit to maintain public order, look out for the security of targets, and protect the substantive autonomy of audiences, even without imposing criminal sanctions. Criminal sanctions could then be reserved for more serious cases involving the use of threatening words or behavior to stir up hatred, or to incite discrimination or violence, against people on the grounds of protected characteristics. This speaks to a division of labor between governments and social media companies that can already be seen in practice, such as when governments impose regulations on social media platforms requiring them to remove illegal hate speech but are silent on the removal of ordinary hate speech, or when social media platforms have separate structures and staff in place for dealing with legal compliance and content moderation respectively.[152]

In fact, in two recent majority decisions—*Moody v. NetChoice, LLC* and *NetChoice, LLC v. Paxton*,[153]—the U.S. Supreme Court has held that when a person's speech is restricted by social media companies, the First Amendment and its various doctrines do not apply, precisely because they are not governments, but private companies. Not only does the First Amendment not apply to private companies like social media platforms, but social media platforms have their own First Amendment rights. For

151. *See* BROWN & SINCLAIR, HATE SPEECH FRONTIERS, *supra* note 4.

152. *Id. See also* ALEXANDER BROWN, MODELS OF GOVERNANCE OF ONLINE HATE SPEECH (2020).

153. Moody v. Netchoice, 603 U.S. 1-30-31 (2024).

example, in his Concurring Opinion in the *NetChoice* cases,[154] Justice Alito upheld these constitutional rights and also pointed to Section 230(c)(2)(A) of Title 47 of the U.S. Code which explicitly grants to providers and users of interactive computer services immunity from liability for "any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected" (the so-called good Samaritan clause).

V. THE CONCEPT OF STIRRING UP

At heart, the above reforms involve shifting the focus of the England and Wales stirring up hatred offenses away from a fixation on hatred and towards a concern with more general conduct of stirring up extreme negative emotions, sentiments, or attitudes toward other persons or groups of people on the grounds of protected characteristics. In this part, I turn to grapple with problematic aspects of the concept of stirring up. These aspects include the problems of defining the phrase "stir up" and of identifying values of sufficient weight or gravity to justify enacting laws that cover not only the stirring up of hatred but also the stirring up of any extreme negative emotions, sentiments, or attitudes.

To get the ball rolling, it is worth reflecting on the legal origins of the phrase 'stir up' in the current England and Wales stirring up hatred offenses. This phrase is present in both the relevant parts of the Public Order Act 1986 and the earlier incarnation of these offenses articulated in Section 6 of the Race Relations Act 1965. Moreover, scholars have suggested that all incitement to hatred laws, and, more generally, group defamation laws, have ancestry in the law of seditious libel.[155] Seditious libel was an English common law offense—abolished by Section 73 of the Coroners and Justice Act 2009—which, in the 1947 case, *R. v. Caunt*,[156] Justice Birkett defined as follows: "a man publishes a seditious libel if he does so with the intention of promoting violence by stirring up hostility and ill-will between different classes of His Majesty's subjects."[157] *Caunt* concerned the publication in a local newspaper of an

154. *Id.* Justice Alito, Concurring Opinion.

155. *See* David Riesman, *Democracy and Defamation: Control of Group Libel*, 42 COLUM. L. REV. 727, 742 (1942); Kenneth Lasson, *Racism in Great Britain: Drawing the Line on Free Speech*, 7 B.C. THIRD WORLD L.J. 161, 162 (1987); and BROWN, HATE SPEECH LAW, *supra* note 2, at 20.

156. R. v. Caunt (1947) 64 L.Q.R. 203 (UK).

157. *Id.*

article containing this sentence: "There is a growing feeling that Britain is in the grip of the Jews [. . .] [and] violence may be the only way to bring them to a sense of their responsibility to the country in which they live." Two much older cases include: *R. v. Osborne*,[158] involving an article that made accusations of child murder against Portuguese Jews living in London in a social context where such libels were likely, in the view of the Court, "to raise tumults and disorders among the people";[159] and *R. v. Burns*,[160] involving a defendant who gave a speech in Trafalgar Square before a crowd of unemployed workers, after which the crowd followed him through the streets of London creating a public disturbance and breaking windows along the way, and which the Court held to be seditious libel by virtue of "stirring up jealousies, hatred and ill-will between different classes of Her Majesty's subjects."[161]

The phrases "stir up" and "stirring up" are ambiguous; it is possible to put different glosses on their meaning. But it is worth noting that part of the central purpose that parliamentarians have often ascribed to the England and Wales stirring up hatred offenses has been to cover cases where a speaker is addressing not members of the group against which hatred is stirred up but an audience of like-minded or at least interested or persuadable persons in whom hatred is intended to be stirred up or likely to be stirred up. This purpose has been cited in response to critics of the relevant stirring up hatred offenses who suggest they might be redundant given that other public order offenses criminalize somewhat similar conduct and can be charged as racially aggravated offenses. (As mentioned above, Section 4A of the Public Order Act 1986 already makes it an offense for a person to use threatening, abusive, or insulting words or behavior with intent to cause another person harassment, alarm or distress and thereby causing that other person harassment, alarm or distress.) For example, during the 2004-2005 parliamentary session the government sought to introduce new legislation to extend the relevant stirring up hatred offenses to cover religious hatred. In this context, then Parliamentary Under-Secretary of State for the Home Office, Caroline Flint MP, offered the following justification for why these offenses are needed:

> Offences under Part I of the Public Order Act 1986 can be religiously aggravated but only catch behaviour that is or is likely to cause harassment,

158. R. v. Osborne (1732) 25 Eng. Rep. 584 (UK).
159. *Id.*
160. R. v. Burns (1732) 148 Eng. Rep. 803 (UK).
161. *Id.*

> alarm or distress to persons likely to see or hear it. These offences do not cover, for example, situations where a person stirs up in his supporters or followers hatred of a group of persons defined by religion where no one from that religion is present to be harassed, alarmed or distressed.[162]

Notwithstanding these points, the meaning of the term stir up is not defined in the Public Order Act 1986, even though the meaning of some other terms *are* defined in the legislation. For example, Section 17 states, "In this Part 'racial hatred' means hatred against a group of persons . . . defined by reference to colour, race, nationality (including citizenship) or ethnic or national origins." Unhelpfully, the Explanatory Notes to the Racial and Religious Hatred Bill 2005 likewise say nothing about the substantive meaning of the term stir up other than to clarify what falls short of stirring up: "legitimate discussion, criticism, or expressions of antipathy or dislike of particular religions or their adherents will not be caught by the offence."[163] By contrast, within the jurisdiction of Scotland, the Information Note accompanying the Hate Crime and Public Order (Scotland) Bill offered the following gloss:

> Stirring up hatred is conduct which encourages others to hate a particular group of people defined by reference to a shared characteristic, for example a racial group. In the context of stirring up hatred, the intention of the perpetrator is that hatred of the group as a whole is aroused in other persons.[164]

Then again, the lack of a statutory definition of the phrase stir up is not unique. The terms "threatening," "abusive," and "insulting" are also undefined in the Public Order Act 1986. Indeed, the statutory definition of the phrase "racial hatred" in the Act does not itself contain a definition of the all-important term "hatred." Moreover, the lack of definition of (some) key terms does not place the stirring up hatred offenses on a different footing to other offenses in the Public Order Act 1986, which similarly lack definitions of some key terms. In law, even in criminal law,

---

162. Joint Committee on Human Rights Eighth Report, Appendix 2a: From Caroline Flint Mp, Parliamentary Under Secretary of State, Home Office, Re Serious Organised Crime & Police Bill, para. 70 (Feb. 3, 2005), https://publications.parliament.uk/pa/jt200405/jtselect/jtrights/60/6013.htm.

163. Explanatory Notes to the Racial and Religious Hatred Bill, *supra* note 77, at para. 27.

164. Information Note Accompanying the Hate Crime and Public Order (Scotland) Bill (n.d.), https://www.gov.scot/binaries/content/documents/govscot/publications/factsheet/2020/04/hate-crime-bill-what-it-will-do/documents/hate-crime-bill-stirring-up-hatred-offences/hate-crime-bill-stirring-up-hatred-offences/govscot%3Adocument/Hate%2BCrime%2BBill%2B-%2BInformation%2BNote%2BPdf%2B-%2BStirring%2BUp%2BHatred%2BOffences%2B-%2BRevised%2BAugust%2B2020.pdf.

there can be virtue in some degree of "open texture" or semantic indeterminacy.[165] Perhaps legislators need to make space or allow flexibility for what society, in the current context or climate, deems to be stirring up racial hatred, for example, especially when it comes to unexpected cases. Maybe a magistrate (summary offenses) or a jury under the guidance of a judge (indictable offenses) are best placed to determine what stirring up racial hatred really means against a wider social backdrop which includes not only the evolving aims and communication strategies of known hate groups but also more general trends in public communication and social controversy such as hyperbole, culture wars, pile-ons, dog whistles, tag lines, political marketing, click bait, impulsive online posting, and so on.

Of course, as well as determining the meaning of "stir up," a magistrate or jury will also need to distinguish between, on the one hand, a person using words or behavior with intent to stir up hatred against another person or group of people and, on the other hand, a person using words or behavior that simply insult, abuse, disparage, vilify, or defame a group of people, or that merely express the speaker's own hatred towards a group of people, or that just put forward, advocate, promote, or defend certain beliefs, ideas, or opinions about a group of people that are not reasonably labelled hatred, or that seek to persuade other people to change their minds or policies towards a group of people.[166]

At any rate, it is no matter that the meaning of stir up is neither defined in the Public Order Act 1986 nor in subsequent acts of Parliament extending its scope to cover other protected characteristics, because the presumption in that instance is for terms to be given their plain or ordinary meaning in the English language. (In the landmark case *Brutus v. Cozens*,[167] the Law Lords established a precedent in relation to public order offenses whereby there is a presumption in favor of interpreting words in statutes as conveying ordinary meanings, unless the legislation or context requires otherwise.) Indeed, this is how the courts have interpreted other key terms used to specify the stirring up hatred offenses in the Public Order Act 1986. For example, the offense of stirring up racial hatred includes the words "threatening, abusive, or insulting," and in a 2005 case, the High Court ruled that these words should be given their

165. Shuangling Li, *A Corpus-Based Study of Vague Language in Legislative Texts: Strategic Use of Vague Terms*, 45 ENGL. SPEC. PURPOSES 98 (2017).

166. *See also* Loren P. Beth, *Group Libel and Free Speech*, 38 MINN L. REV. 167, 178 (1955); and Twomey, *supra* note 1, at 243.

167. Brutus v. Cozens, [1972] 2 ER (Eng. Rep.) 1297 (UK).

ordinary meaning.[168] Likewise, in a 2010 case, a Crown Court held that the statutory term "threatening" should carry its ordinary meaning, other things remaining equal.[169] The test is this: Would a reasonable member of the public consider the given words or behavior to be threatening, abusive, or insulting?(Interestingly, within the jurisdiction of Scotland, Part 3 of the Hate Crime and Public Order (Scotland) Act 2021 explicitly states (within the statute itself) that the offence covers a person who "behaves in a manner that a reasonable person would consider to be threatening, abusive or insulting, or [. . .] communicates to another person material that a reasonable person would consider to be threatening, abusive or insulting".) Moreover, given the lack of statutory definition of the term "hatred" in the Public Order Act 1986, as Anne Twomey points out, "[i]t is therefore left to the jury to determine whether 'hatred' means intense detestation or merely ill-will or dislike."[170] Furthermore, in some countries, courts have explicitly applied the ordinary meaning approach to interpreting phrases similar to stir up, specifically the word "incite," within relevant hate speech laws.[171] What is more, within the jurisdiction of Scotland, the relevant legislation explicitly states that a person can commit an offense if "a reasonable person would consider the behavior or the communication of the material to be likely to result in hatred being stirred up against such a group."[172]

Therefore, I propose that within England and Wales the corresponding test for the phrase "stir up" in the relevant offenses should be this: Would a reasonable member of the public consider a person's use of given words or behavior to amount to stirring up hatred? In coming to an understanding of the ordinary meaning of the phrase, a reasonable person might also consult the meaning of related or close cousin terms. Talk of "stirring up" enmity between social groups and/or hatred toward a particular group of people certainly has semantic affinities with the

168. Director of Public Prosecutions v. Humphrey, EWHC 822 (Admin) [2005].

169. *R. v. Bamber*, No. T20091255 [2010] (Preston Crown Court, June 2010) (Eng.).

170. Twomey, *supra* note 1, at 242.

171. For example, in *Burns v. Dye*, NSWADT 32 [2002], the NSW Administrative Decisions Tribunal held that, with regards to s. 49ZT(1) of the Anti-Discrimination Act 1977 (NSW), according to which it is unlawful to incite hatred of homosexuals: "The word 'incite' is to be given its ordinary natural meaning which is to 'urge, spur on . . . stir up, animate; stimulate to do something' (New Shorter Oxford English Dictionary, 1993) (Oxford); 'urge on; stimulate or prompt to action' (the Macquarie Dictionary, third edition, 1997) (Macquarie)." *Id.* at para. 19. A similar approach was endorsed in *Kimble and Souris v. Orr*, NSWADT 49 [2003], in relation to interpreting s. 20C(1) of the Act: "The word 'incite' should be given its ordinary English meaning, namely, to urge, spur on, stir up, animate, stimulate, or prompt action. It is not sufficient if the words merely convey hatred or express serious contempt or severe ridicule." *Id.* at para. 62.

172. Part 3 of the Hate Crime and Public Order (Scotland) Act 2021.

English language idioms "stirring the pot" and "shit-stirring," which similarly speak to deliberately causing trouble or controversy, such as by agitating other people so as to cause a reaction or conflict. Magistrates and jurors might also find their way to understanding what stir up means by reflecting on the meaning of similar terms or synonyms such as "promote," "incite," "foment," "provoke," or "evince," for example.

In practice, the act of stirring up hatred, or stirring up any extreme negative emotions, sentiments, or attitudes, could elicit many different types of changes in audiences' emotional and/or cognitive states. To stir up hatred, for example, could be to do one or more of the following: to enliven, build, increase, whip up, or grow hatred, as in, turn weaker antipathy towards a targeted group into something much stronger or more intense; to arouse, kindle, awaken, animate, or ignite previously nascent, dormant, or unconscious hatred; to trigger, activate, or switch on hatred in persons who already have a pattern, habit, or disposition toward hatred; to socialize or internalize hatred, such as to transform or evolve single episodes of hatred into something like a pattern, habit, or disposition toward hatred; to normalize, legitimate, justify, excuse, or make people feel more comfortable about, entitled to, or emboldened in the hatred they already harbor; to channel or provide a focal point for existing hatred, such as by suggesting that a general class of persons who deserve hatred contain a subset who are especially deserving of hatred; or to harness or redirect hatred towards new targets, such as by pointing out that members of the targeted group are in fact members of a general class of persons who therefore also deserve hatred; or to create or spark hatred ex nihilo, where previously no ill-will or hostility existed whatsoever.[173]

It will be up to magistrates (summary offenses) or juries under the guidance of judges (indictable offenses) to figure out what sort of stirring up of hatred might have been intended and likely in given cases, if any at all. To take one example, in *R. v. Hutchinson*,[174] the defendant was found guilty of stirring up racial hatred under the Public Order Act 1986 in relation to posting racist memes and comments on VK, a Russian social media site. This included posting that he was "waiting for my white race to wake up and fight back," that he was "looking for 40 [white] men" to join his campaign against Black people. Each magistrate or juror might

---

173. David Kretzmer, *Freedom of Speech and Racism*, 8 CARDOZO L. REV. 445, 464 (1987); Brown, *The Racial and Religious Hatred Act 2006*, *supra* note 1, at 6; Rosenfeld, *supra* note 70, at 259; and RAE LANGTON, *Beyond Belief: Pragmatics in Hate Speech and Pornography*, in SPEECH AND HARM: CONTROVERSIES OVER FREE SPEECH (I. Maitra & M. McGowan eds., 2012), at 89.

174. R. v. Hutchinson [2022] (Kingston Criminal Court, Oct. 18, 2022) (Eng.).

put their own gloss on what it means to stir up racial hatred as applied to the facts of this case. Each magistrate or juror might have a different sense in mind, whether that is growing, sparking, channeling, harnessing, or triggering hatred. But this variation in itself need not be problematic. Just as "hate speech" is a "family resemblance" term in the Wittgensteinian sense,[175] so might be "stir up." This means speakers can understand and competently use the phrase without having a precise definition in mind.

To clarify, I do not mean to suggest that applying the phrase "stir up" will always be easy. For example, a jury in a crown court case might be asked to come to a verdict on a political leader whose use of language and declared intentions in using that language are precisely calculated to stay on the right side of the law, thus making it difficult to say for sure whether in fact it is a case of stirring up hatred. The particular speech context and wider information about the official roles and affiliations of the speaker and what the speaker has said or posted previously might be useful in building a picture, but might not be decisive. In a case from 2006, a jury at Leeds Crown Court were informed that Mark Collett, the then chairman of the British National Party (BNP), a far-right party, had made a speech at a public house to a group of party members and others, in which he allegedly stated:

> I honestly don't hate asylum seekers—these people are cockroaches and they're doing what cockroaches do because cockroaches can't help what they do, they just do it, like cats meow and dogs bark. They do it because they are what they are and they'll do what they do. The people I hate are the white politicians who have sold us down the line. I'd rather die today with my pride intact, fighting for what I believe in, than live the rest of my life as a sniffling pathetic slave to a multicultural society. This is our battle for Britain.[176]

The jury was asked to decide whether the prosecution had proved beyond a reasonable doubt that Collett had used threatening, abusive, or insulting words or behavior, and either intended thereby to stir up racial hatred or having regard to all the circumstances racial hatred was likely to be stirred up thereby. The jury found Collett not guilty.[177] One can only conjecture as to the reasons for the not guilty verdict, but perhaps jurors thought it possible that Collett was not intending to stir up hatred.

---

175. *See* Brown, *What Is Hate Speech? Part 2*, *supra* note 126.

176. Staff and agencies, *Jury Hears of BNP's 'Multiracial Hell' Speech*, THE GUARDIAN (Nov. 3, 2006), https://www.theguardian.com/uk/2006/nov/03/ukcrime.thefarright.

177. Staff and agencies, *BNP Leader Cleared of Race Hate Charges*, THE GUARDIAN (Nov. 10, 2006), https://www.theguardian.com/politics/2006/nov/10/thefarright.uk.

Moreover, jurors may have taken a view that, given the nature of the audience, no hatred would be likely stirred up that was not already present (i.e. that the audience was already corrupted).[178] Some of the jurors might have asked themselves and been unable to answer the following question: In what sense can it be said that a chairman of the BNP was likely to stir up hatred among a group of his own supporters given that their very attendance of the meeting might be partly explained by their pre-existing hatred of asylum seekers? As Geoffrey Bindman puts it, "the hatred may already be felt by the audience."[179] Then again, it is surely incumbent upon the prosecution, the judge, and even other jurors to point out to those who would make an assumption of this kind that the phrase "stir up" can have multiple meanings. They might suggest to jurors that stirring up hatred can mean increasing, whipping up, or growing hatred already felt by the audience; or it could mean tuning into hatred already felt for people of color and redirecting it towards asylum seekers as a social group.

Interestingly, Section 6(2) of the earlier Race Relations Act 1965 included an exemption to the offense of stirring up racial hatred in cases where a person publishes or distributes written matter to a section of the public "consisting exclusively of members of an association of which the person publishing or distributing is a member."[180] However, this exemption was omitted from the Public Order Act 1986. In the white paper that preceded the latter, the government explained its rationale for the omission as follows:

> This provision was intended to protect freedom of expression within a group holding particular views, but it is possible that even those who already hold racialist views may be incited or incited further to racial hatred [. . .]; accordingly it proposes to remove the exemption for material circulated to members of an association.[181]

The government's logic seems to be that stirring up hatred can take many forms and the mere fact that a speaker is "preaching to the choir" does not necessarily mean that the stirring up of hatred is unable or unlikely to take place. So, in the case of written material circulated among members or followers of a far-right party or movement, just as with words delivered orally in a speech to similar people while assembled at a meeting, march, or rally, it might be unlikely that the stirring up primarily takes the form of the speaker creating hatred of a targeted group ex nihilo,

178. Neller, *supra* note 45, at 126.
179. Bindman, *Outlawing Hate Speech*, *supra* note 1, at 18.
180. § 6(2) of the earlier Race Relations Act 1965.
181. REVIEW OF PUBLIC ORDER LAW, 1985 CMND. 9510, 39 (May 16, 1985).

where previously no hostility existed whatsoever. Nevertheless, this does not exhaust the forms of stirring up hatred. Perhaps the language could be used to trigger hatred that was previously dormant, to make people feel more relaxed about the hatred they already feel such that they lean into it further, or else to turn or consolidate occasional hatred into a pattern, habit, or disposition.

At any rate, the key point is that, under my proposals for a model stirring up offense, the term "stir up" is to retain its plain or ordinary meaning as per the existing England and Wales stirring up hatred offenses. This means I do not propose to give the term a narrower, statutory definition. However, this is not without implications. Arguably, retaining the ordinary meanings of terms in statutes can increase the risk of them being overbroad. This is because the ordinary meaning of natural language terms can, although not always, involve polysemy, as in, the coexistence of many possible meanings of the same term. It is clear from what I have said above that the phrase "stir up" is polysemous. Interestingly, exactly this point about the overbreadth risk associated with plain meanings was emphasized by the U.S. Supreme Court in *U.S. v. Hansen* in relation to the terms "encourage" and "induce."

> To judge whether a statute is overbroad, we must first determine what it covers. Recall that §1324(a)(1)(A)(iv) makes it unlawful to "encourag[e] or induc[e] an alien to come to, enter, or reside in the United States, knowing or in reckless disregard of the fact that such coming to, entry, or residence is or will be in violation of law." The issue is whether Congress used "encourage" and "induce" as terms of art referring to criminal solicitation and facilitation (thus capturing only a narrow band of speech) or instead as those terms are used in everyday conversation (thus encompassing a broader swath). An overbreadth challenge obviously has better odds on the latter view.[182]

This illustrates once again the tension, or push and pull, between the viewpoint discrimination and overbreadth doctrines. Put simply, I have proposed that the England and Wales stirring up hatred offenses be reformed to push them away from an overemphasis on the hatred element and more towards the stirring up element in order to mitigate viewpoint discrimination, but if the stirring up element depends on an ordinary meaning approach to statute interpretation, then this can pull the offense closer towards overbreadth.

Nevertheless, as made clear in *Hansen*, just because a law sweeps up some speech that ought to be protected does not mean that the law is

182. U.S. v. Hansen, 599 U.S. 5 (2023).

unconstitutionally overbroad. There must be a realistic prospect that the ordinary meaning of the term will mean the law sweeps up protected speech and, what is more, that the amount of protected speech that is swept up "must be substantially disproportionate to the statute's lawful sweep."[183] I would argue that because of other elements within my model stirring up offense, most notably that a prosecution must show both intent to stir up and that actual stirring up is likely, then the mere fact that the term "stir up" could capture a wide swath of activities (as listed above) does not mean the model offense would be unconstitutionally overbroad.

Notwithstanding these points, there is a deeper question that needs to be answered about stirring up: What is it about stirring up that renders it proscribable conduct? My answer once again stresses threats to substantive autonomy and, specifically, the way stirring up hatred can circumvent or curtail processes of independent, rational deliberation in receivers. That said, I do not mean to suggest the value of substantive autonomy is the only reason. Protecting public order is an important reason often highlighted by parliamentarians.[184] Another reason is combating the climate of hatred and fear, and the associated heightened risks to victims' safety and sense of personal safety, to which the stirring up of hatred contributes.[185] But autonomy is a reason that hitherto has not received the attention it deserves from parliamentarians.

Exactly how can words or behavior stir up extreme negative emotions, sentiments, or attitudes in receivers, and in what way might these mechanisms pose a threat to receivers' substantive autonomy? To focus on common types of hate speech, perhaps negative stereotypes, for instance, can stir up extreme attitudes at an unconscious level. When some people read or hear stereotypes about minority groups, it might confirm what they already unconsciously believed about them. This confirmation of unconscious beliefs could be a potent mechanism, capable of bringing to the surface, triggering, or activating other extreme negative emotions, sentiments, or attitudes in automatic, unreflective, and non-deliberative ways. Indeed, in some instances, a person with unconscious bias might read or listen to a negative stereotype and this activates a prejudiced attitude toward the targets of the speech, but this attitude remains at an unconscious level along with the unconscious bias. The fact that all of this occurs at an unconscious level could account for why a person might

183. *Id.*

184. BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3, at ch. 3; Brown, *The "Who?" Question in the Hate Speech Debate: Part 2*, *supra* note 1, at 26-33; Neller, *supra* note 45.

185. BROWN, HATE SPEECH LAW, *supra* note 2, at 66-75.

protest they are not prejudiced against the targets. Moreover, protesting they are not prejudiced could be an unconscious psychological defense mechanism against cognitive dissonance.

But how about mechanisms that operate at a conscious level that people are directly aware of and experience or feel? In his analysis of terms such as "promoting," "inciting," and "stirring up" that are present in incitement to hatred laws not only in the U.K. but also in Canada and elsewhere,[186] W. L. Sumner suggests the following semantic commonality: "appealing to the passions rather than to reason."[187] "[I]t works through getting the subject worked up or agitated rather than by offering a convincing argument" and "contrasts with counselling, or advising, or persuading."[188] A similar analysis has been proffered by the Hungarian Constitutional Court:

> According to the law, the term "incitement" is not the expression of some unfavourable and offensive opinion, but virulent outbursts which are capable of whipping up intense emotions in the majority of people which, upon giving rise to hatred, may result in disturbing the social order and peace [. . .]. This way, criticism, disapproval, objections or even offensive declarations do not constitute incitement; incitement occurs only when the expressions, comments etc. do not address reason but they seek to influence the world of emotions and are capable of arousing passion and hostile feelings.[189]

To offer my own take: It is possible that sometimes the act of stirring up hatred operates at the level of emotions all the way down the line, such as when a person uses highly emotive language to trigger an emotional response in the audience, thereby circumventing or curtailing the latter's processes of independent, rational deliberation. Perhaps racial slurs or epithets work in this way in certain cases. Consider a person who uses a racial slur to express their own anger, resentment, or hatred toward the targets of the slur, thereby arousing a similar level of anger, resentment, or hatred in the audience, as a kind of emotional call and response. Maybe threatening words or behavior also stir up extreme negative emotions or sentiments as a response to the emotions they express. The threatening words or behavior express anger and the audience's response is fear. In

---

186. *Id.* at 26-27.

187. L. W. SUMNER, *Incitement and the Regulation of Hate Speech in Canada: A Philosophical Analysis*, *in* EXTREME SPEECH AND DEMOCRACY 215 (I. Hare & J. Weinstein eds., 2009).

188. *Id.*

189. Decision 12/1999 (V. 21.) AB (Budapest, May 19, 1999).

the grips of these extreme negative emotions, audiences are less able to rationally deliberate about what to think about any ideas or opinions encoded in this language and about what to do or how to react.

No doubt, in other cases, the stirring up of extreme negative emotions, sentiments, or attitudes works at both cognitive and non-cognitive (emotional) levels simultaneously, and in mutually reinforcing ways. Suppose a person takes to social media to portray transgender people as "deceptive and dangerous sex predators," to make threatening comments about what people should do to any transgender person who "dares to enter" a restroom with their wife or daughter, and to ask readers to click "like" or "share" if they feel the same way. The social media post might encode a negative stereotype and a misleading or false narrative about transgender people and express anger towards them. The instinctive response among some people reading the post might be to agree with the stereotype and the narrative, to automatically conclude that transgender people deserve the anger and the implied violence, and to end up having their own prejudices activated and their own anger whipped up. There will be a mixture of cognitive and non-cognitive elements that work together here to circumvent or curtail processes of rational deliberation.[190]

VI. SENTENCING GUIDELINES

A final piece of the jigsaw in reforming the current England and Wales stirring up hatred offenses concerns the sentences handed down by judges to persons who plead guilty or are found guilty. The Public Order Act 1986 sets out a range of sentences that magistrates and judges have available to them (including custodial sentences and community orders), and in deciding the length of sentences in given cases they consider a variety of different factors. This often overlooked stage of the legal process again implicates issues of substantive autonomy and ought to be reformed accordingly.

The Coroners and Justice Act 2009 gives the Sentencing Council for England and Wales a statutory duty to issue guidelines on sentencing within that particular jurisdiction, which the courts must follow unless it is in the interests of justice not to do so.[191] This is an independent, non-departmental public body—or quango—that aims to promote greater

190. Another relevant factor here is the way social media encourages instantaneous, impulsive engagement with content, including hate speech content. *See* Alexander Brown, *What Is So Special About Online (as Compared to Offline) Hate Speech?*, 18 ETHNICITIES 297 (2018); and Alexander Brown, *The Internet of Hate: Comparing the Nature, Harms, and Regulatory Challenges of Online and Offline Hate Speech*, 53 GA J. INT'L & COMP. L. (2025).

191. Coroners and Justice Act 2009.

transparency and consistency in sentencing, thereby increasing public awareness of, and confidence in, how the criminal justice system works, while at the same time operating at arm's length from the government of the day, and also seeking to maintain the independence of the judiciary. The guidelines on sentencing issued by the Sentencing Council in relation to the existing England and Wales stirring up hatred offenses identify various aggravating factors.[192] It is noticeable how these factors implicitly speak to different rationales behind the legislation itself. For example, in relation to the rationales of maintaining public order and safeguarding targets, the guidelines state that a person convicted of stirring up hatred offenses—the elements of which already include intention to stir up hatred—has greater culpability if they also had "intention to incite serious violence." They also state that "the level of harm that has been caused or was intended to be caused to the victim" is increased if the speech or behavior "directly encourages activity which threatens or endangers life" or if there is "[w]idespread dissemination" of the relevant statement, publication, or performance.[193] However, I believe that other aggravating factors set forth in the guidelines implicate not only these rationales but also the protection of substantive autonomy.

In particular, the guidelines state that a person convicted of stirring up hatred offenses has greater culpability if the offender uses (or abuses) a "position of trust, authority or influence to stir up hatred."[194] It also identifies as another aggravating factor that an audience that is "vulnerable/impressionable."[195] I believe these sentencing guidelines can be motivated not merely by the rationales of maintaining public order and safeguarding targets but also by the aim of protecting the substantive autonomy of receivers. The key point is that in some instances the stirring up of hatred can constitute the exercise of "undue influence" over the receiver,[196] in the sense that it can inhibit the receiver from deciding for themselves what to think about, and how to behave towards, the person or group of people hatred is being stirred up against, such as by circumventing or curtailing the receiver's rational deliberation and, in so doing, effectively substituting the receiver's thinking and decision-making (or will) with that of the speaker. I believe these same guidelines

192. Sentencing Council, *Aggravating and Mitigating Factors*, Sentencing Council (Apr. 19, 2025), https://www.sentencingcouncil.org.uk/explanatory-material/magistrates-court/item/aggravating-and-mitigating-factors/.

193. *Id.*

194. *Id.*

195. *Id.*

196. *See* BROWN, HATE SPEECH LAW, *supra* note 2, at 60.

can, and should, apply to the model stirring up offense I have proposed, which most notably broadens the focus beyond stirring up hatred to stirring up any extreme negativity towards other people on the grounds of protected characteristics.

In applying these guidelines, the question is not simply whether a speaker has the linguistic tools and skills necessary to stir up extreme negative emotions, sentiments, or attitudes in an audience. It is also about whether they did so relying on a "position of trust, authority or influence," without which they might not have been able to stir up hate or would have been less successful doing so, such as if the position enables them to use coercion on the audience or to circumvent or curtail the receiver's independent, rational deliberation.[197] In addition, it is relevant whether the speaker intended and was likely to stir up extreme negativity in a "vulnerable/impressionable audience," to whom they owed a duty of care or special epistemic responsibility.[198]

My recommendation for needed reform in this area is that the sentencing guidelines be amended to provide some concrete examples of the sorts of positions of trust, authority, or influence, and the kinds of vulnerable/impressionable audiences covered, provided that the guidelines make clear that these are merely illustrations and are not meant as exhaustive lists. I would argue that providing some concrete examples could help not only to demystify sentencing for victims and the public alike but also to give persons greater ability to anticipate the particular legal consequences of their actions, so individuals could make decisions about their behavior based on foresight of both what is likely to constitute an offense and what judges are likely to deem high culpability or aggravated factors if they plead guilty or else plead not guilty but are nevertheless found guilty by a magistrate or jury. Arguably, these features could increase public confidence in the criminal justice system, while also enhancing a key aspect of the rule of law, namely predictability.

I end this part by offering two concrete examples that I propose should be added to the existing sentencing guidelines. The first is of a position of trust, authority, or influence that might give the speaker the ability to coerce the audience, or else circumvent or curtail the receiver's normal processes of independent, rational deliberation. It is the example of political leaders, campaigners, activists, or candidates using threatening words or behavior to stir up extreme negative emotions, sentiments, or attitudes among their members, followers, or other audiences who they

197. *Id.* at 61-64.
198. *Id.* at 64-66.

might persuade due to their position of trust, authority, or influence. There is something special about the positions occupied by such people that means that *some* people may be more willing, at an epistemic as well as emotional level, to listen to, to give credence to, to accept without further investigation, to regard as normal and legitimate, and so on, what they say, and that *some* people may be more susceptible to being emotionally affected (or manipulated) by what they say. These mechanisms, along with facts about the size of the audience, might justify treating individuals in political positions of trust, authority or influence as an aggravating factor at sentencing.[199]

Consider *R. v. Fielding-Morriss*.[200] In this case a jury found a parliamentary candidate guilty of three counts of stirring up racial hatred based on a series of blog posts she had published online as part of her campaign. In the posts she had praised Hitler, advocated the building of "new and better death camps" for Jews, likened Jewish asylum seekers to "termites," and called for Britain to be "white only." The jury found her guilty of the offenses, which were committed during a period she twice stood as a candidate to be an MP (winning 137 votes in a by-election in February 2017 and then 210 votes in the general election in June 2017) and was the leader, albeit also the only member, of a registered political party (the Abolish Magna Carta Reinstate Monarchy Party). In his sentencing remarks, Recorder (Judge) Taylor, cited the defendant's political candidacy as an aggravating factor.

> The background to this case is that you stood as a parliamentary candidate. Your manifesto, which was published on a website and in a blog, contained material that formed the subject of the three counts on the indictment. [. . .] The fact of the matter is that you intended to stir up racial hatred. The fact you were standing in a general election as a parliamentary candidate aggravates this case, because you were putting views forward to an electorate.[201]

It is not entirely clear from these remarks why Recorder (Judge) Taylor deemed "putting views forward to an electorate" an aggravating factor. One factor could be the potential for heightened media exposure and audience size. But there is also the factor that being a parliamentary

---

199. *See also* BROWN & SINCLAIR, THE POLITICS OF HATE SPEECH LAWS, *supra* note 3, at 430-31.

200. R. v. Fielding-Morriss, No. T20170342 [2018] (Stoke-on-Trent Crown Court, Oct. 5, 2018) (Eng.).

201. Reporter (Judge) Taylor's sentencing remarks (Oct. 5, 2018), www.thelawpages.com/court-cases/Barbara-Fielding-Morriss-23882-1.law.

candidate is a position of trust and influence, in the sense that *some* people might be more likely to accept her words as epistemically authoritative and/or legitimate because they were spoken by a candidate, fallacies that could make the words more potent than they would otherwise be.

My second example is of a vulnerable or impressionable audience to whom a speaker might owe a duty of care or special epistemic responsibility. Consider religious leaders or teachers using threatening words or behavior to stir up extreme negative emotions, sentiments, or attitudes among their congregation, followers, or students, against other people on the grounds of protected characteristics. It is possible that due to the status of the audience, or the relationship between speaker and audience, in these cases, the audience may be more prone to trust the source, more prone to suggestibility, and more prone to agreeableness or compliance. This form of stirring up is aggravated because the speaker might owe a duty of care or special epistemic responsibility to congregations, followers, or students precisely because of the latter's vulnerability or impressionability. Stirring up hatred among children (which legally speaking in the U.K. means those under the age of 18) might be an especially serious (egregious) case of stirring up hatred among the vulnerable or impressionable. When the audience is children, as well as having less formal power to decline, object to, or avoid the religious instruction given to them, they could be, on average, less equipped to undertake independent, rational deliberation in response to what the speaker is telling them, including a religious leader or teacher who is seeking to incite or persuade them to adopt an extreme negative emotion, sentiment, or attitude against a certain group. They might also be more prone to imitation.[202]

Interestingly, Robert Simpson has argued that one could accept in general terms that the state impinges upon people's formal autonomy by trying to control what ideas they are exposed to, while at the same time regarding children as a special case, such that their substantive autonomy is safeguarded by the state trying to control what ideas they are exposed to, including when those ideas are hateful.[203] I believe exactly the same

---

202. On the point about imitation, *see* S.J. Ceci & R.D. Friedman, *The Suggestibility of Children: Scientific Research and Legal Implications*, 86 CORNELL L. REV. 34 (2000). There may be some similarities here with how exposure to depictions of violence on television can cause people, especially children, to instinctively imitate what they see, and in ways that can bypass autonomous deliberation, and in automatic and unconscious ways. *See* Susan Hurley, *Imitation, Media Violence, and Freedom of Speech*, 117 PHIL. STUD. 165 (2004).

203. Robert Mark Simpson, *"Won't Somebody Please Think of the Children?" Hate Speech, Harm, and Childhood*, 38 L. & PHIL. 79, 106 (2019).

point holds when the case involves not exposure to ideas simpliciter but exposure to the actions of a person who uses threatening words or behavior with intent to stir up extreme negative emotions, sentiments, or attitudes.

Consider *R. v. El-Faisal*.[204] In this case, the police were investigating possible al-Qaeda links in the U.K. and found tape recordings of speeches given by the Jamaican-born Muslim cleric Abdullah el-Faisal labelled "No peace with the Jews" and "Jewish Traits." He was convicted *inter alia* of two counts of distributing threatening recordings with intent to stir up racial hatred—the first Muslim cleric to be convicted of such offenses in the U.K. For each of the two counts el-Faisal received a sentence of twelve months imprisonment. In his sentencing remarks Judge Beaumont touched upon el-Faisal's special responsibilities.

> In my judgment, your offending was aggravated by the fact that as a cleric you were sent to this country to preach and minister to the Muslim community in London, and so had a responsibility to the young and impressionable within that community at times of conflict abroad and understandable tensions in the communities here over the period which is spanned by the indictment.[205]

Once again, it is uncertain whether Judge Beaumont meant to lump the young and impressionable into a single class. Perhaps it is possible that anyone receiving extensive guidance and pastoral care from someone in a position of religious authority is to some degree impressionable. Either way, what is clear is that Judge Beaumont was of the view that el-Faisal's audience were young and impressionable, and this aggravated the offending.

VII. CONCLUSION

In this Article I critically examined several underexplored issues related to the England and Wales stirring up hatred offenses, including: the extent to which they involve viewpoint discrimination, the prospect that avoiding or mitigating viewpoint discrimination would make the offenses overbroad, whether there are any unique problems in courts applying an ordinary meaning approach to the phrase "stir up," and how to make sense of some of the current sentencing guidelines for these offenses. Along the way, I appealed to the value of substantive autonomy, as well as some other important rationales, such as maintaining public

204. R. v. El-Faisal, No. T20027343 [2023] (Central Criminal Court, Feb. 24, 2003) (Eng.).

205. Judge Beaumont sentencing remarks (Mar. 7, 2003). Transcript obtained from Smith Bernal Reporting Ltd. with consent of Judge Beaumont.

order. I also recommended a number of reforms to the offenses as follows: that the list of protected characteristics be expanded; that for all protected characteristics the offenses should be standardized, in particular to cover only the use of threatening words or behavior, and to include a conjunction of the elements of "intent to stir up hatred" and "likely to stir up hatred"; that the offenses be widened to cover not only the stirring up of hatred but also the stirring up of any extreme negative emotions, sentiments, or attitudes against persons or groups of people based on protected characteristics; that the offenses be enlarged to also cover incitement to discrimination or violence; that the offenses be given extraterritorial application; and that the sentencing guidelines be amended to provide concrete illustrations of a speaker holding a "position of trust, authority, or influence" and of an audience being "vulnerable/ impressionable," specifically the examples of political figures and the congregations, followers, or students of religious leaders or teachers.

These reforms are long overdue. Britain has had statutes banning the stirring up of racial hatred since 1965 (not to mention older statutes covering broader, yet related public order offenses, and the much older common law offense of seditious libel). Of course, it would be unreasonable to judge the success or failure of the stirring up hatred offenses simply by asking whether Britain continues to see speakers stir up hatred against minorities and to witness its streets periodically being taken over by riots indirectly caused by such stirring up. Britain might continue to experience these phenomena despite the laws, but nevertheless society might have been in an even worse position had, contrary to fact, these laws never been passed. Even so, the stirring up hatred offenses have seen many reforms over the decades and continue to be a work in progress. Britain prides itself on its tolerance both of minorities and of speech. But in promoting tolerance of minorities, it has embraced stirring up hatred offenses that discriminate among viewpoints, which in other legal cultures, like the U.S., is seen as inimical to tolerance of speech. My recommendations take this problem seriously and push the offenses toward a greater inclusion of viewpoints. In particular, the reforms move away from a fixation with stirring up hatred to recognize the various extreme negative emotions, sentiments, or attitudes that can be stirred up. To balance out the equally great threat of overbreadth, however, my recommendations also draw the offenses more narrowly on certain key points, such as standardizing the offenses for all protected characteristics with regards to only covering "threatening words or behavior," and reverting the law back to its 1965 incarnation by requiring both of the phrases "intent to stir up hatred" and "likely to stir up hatred."

In the slow march of progress, space must be made for aspects that the legislation used to get right. This is the non-linear nature of legislative evolution.

To return to the case of Tommy Robinson mentioned in Part II of this Article, I believe my model stirring up offense would provide prosecutors with a finely balanced tool to combat some but not all of his Islamophobic posts and videos (especially if the model offense were given extraterritorial applicability). On the one hand, prosecutors need only demonstrate words were intended and likely to stir up an extreme negative emotion, sentiment, or attitude on grounds of religion, whether hatred, contempt, grievance, resentment, revenge, anger, fear, abhorrence, scorn, derision, superiority, bigotry, prejudice, or similar. Prosecutors would *not* have to try to fit words into a narrow rubric of stirring up religious hatred. On the other hand, prosecutors would still need to show the words were threatening for the offense to be applicable. Looking again at the illustrations of Robinson's content cited in Part II of this Article, the speech and YouTube post from 2011 arguably involved Robinson using threatening words against Muslims deliberately to stir up a range of extreme negative emotions, sentiments, and attitudes (and potentially even to incite discrimination or violence), including to stir up fear in Muslims and also to stir up resentment, anger, and revenge in non-Muslims, and so this sort of post would be covered by my model stirring up offense. By contrast, the three X posts from 2024 responding to the Southport stabbings arguably involved defamation of religion and potentially group defamation, and also the stirring up of anger towards Muslims (and the government) but *without* using threatening words and, therefore, these sorts of posts would *not* be covered by my model stirring up offense. This is to be contrasted with the words used by the first person to be convicted of stirring up hatred offenses in relation to the aftermath of the Southport stabbings—and the sorts of words that would be covered by my model stirring up offense. Specifically, in *R. v. Parlour*,[206] the defendant was found guilty of stirring up racial hatred in accordance with Part 3 of the Public Order Act 1986 for posting on Facebook *inter alia*:

> Every man and their dog should be smashing fuck out Britannia Hotel. [. . .] Because their over here, given life of reilly [sic] off the tax us hard working people earn, when it could be put to better use. Come over here with no work visa, no trade to their name and sit down and doss and then there's more people being put out homeless each year, they get top band priority on housing and many more other reasons.

206. R. v. Parlour [2024] (Leeds Magistrates Court, Aug. 6, 2024) (Eng.).

In his sentencing remarks, Judge Guy Kearl noted:

> For the offence of publishing written material in order to stir up racial hatred there are sentencing guidelines which I must and will follow. [. . .] Your position is aggravated by the timing of your post, namely that it was at a time of social unrest and particular sensitivity across the country.[207]

Just as important could be the symbolic, dialectical, and political effect of my model stirring up offense. First, because the offense would cover other protected characteristics besides race and religion, Robinson could not reasonably protest that the law has been drafted simply to protect one or two minorities (e.g. asylum seekers, Muslims) at the expense of the freedom of "ordinary people." The law could end up protecting vast swaths of the population, depending on how long the list of protected characteristics becomes. Second, *if* political affiliation and beliefs were also included as a protected characteristic, then Robinson could not reasonably protest that his stirrings up against Muslims were being banned but not allegedly equivalent stirrings up against *him* on grounds of his political affiliation and beliefs. But, then again, if political affiliation and beliefs were *not* included, then Robinson would be free to stir up hatred against religious liberals just as religious liberals could stir up hatred against him. So, either way, he could not justly complain he was being ruled with an uneven hand. Third, my proposed reform to sentencing guidelines whereby political figures are cited as illustrations of the aggravating factor of being in a position of trust, authority, or influence, mean Robinson could not reasonably protest that liability to a more severe punishment was unpredictable.

I end by asking a question about broader relevance. How relevant are my arguments to other sorts of stirring up or promoting of extreme negative emotions, sentiments, or attitudes besides the stirring up of extreme negativity *against* people on the grounds of protected characteristics? No doubt my arguments could be applicable to many forms of toxic material circulating online and offline. One obvious example might be material that normalizes, glorifies, or promotes suicide or anorexia, for example. This might be especially serious (egregious) when the material is aimed at children and/or when the speaker knows or can be reasonably expected to know that the people will be suggested the material by social media algorithms or will find their way to the material precisely by having a track record of looking for such material or by identifying as having suicidal thoughts or an eating disorder. The practical

---

207. Judge Guy Kearl sentencing remarks (Aug. 9, 2024), https://www.judiciary.uk/wp-content/uploads/2024/08/Jordan-Parlour.-Media-Posts.-Final.pdf.

question is whether existing legislation in this area is already sufficient or also needs to be reformed along similar lines to the reforms I have proposed to the stirring up hatred offenses. For example, currently within England and Wales, a person commits an offense under Section 2 of the Suicide Act 1961 if they perform an "an act capable of encouraging or assisting the suicide or attempted suicide of another person" and the "act was intended to encourage or assist suicide or an attempt at suicide."[208] It is an open question whether this law also needs reform, along similar lines to the reforms I have suggested for the stirring up hatred offenses, so as to address issues of viewpoint discrimination and overbreadth. The underlying rationales for the legislation might include preventing suicide but also protecting the receiver's substantive autonomy. Likewise, currently within the U.K., a person commits an offense under Section 184 of the Online Safety Act 2023 if, with intent, they perform an act "capable of encouraging" the "serious self-harm" of another person.[209] This offense is applicable to encouraging another person to engage in harmful weight loss such as by encouraging prolonged nutritional deprivation that exposes them to risk of serious ill-health or death. Once again, it is open to debate whether this legislation strikes the right balance between the viewpoint discrimination and overbreadth doctrines. Here, too, the underlying rationales for the legislation could include promoting public health but also protecting the receiver's substantive autonomy.

208. § 2 of the Suicide Act 1961.
209. § 184 of the Online Safety Act 2023.