The Expression of Hate in Hate Speech

TERESA MARQUES (ORCID)

ABSTRACT *In this article, I argue that hate speech expresses hate, and I answer some objections to expressivist views. First, I briefly comment on some limitations of pragmatic accounts of harmful speech. I then present an expressive-normative view of derogatory discourse according to which it is expressive of an affective state by presupposing it. A linguistic act expressive of an affective state inherits the normativity that is constitutive of that state, as directed to its intentional object. If the act is successful, it updates the conversational context with the normative appraisal conditions of the affective state presupposed. I argue that this model can be applied to hate speech. I rely on current research on the psychology of hate to identify the appraisal conditions, action tendencies, and motivational goals characteristic of hate. The account supplements other pragmatic accounts of hate speech.*

1. Introduction

Currently, the view that hate speech expresses hate is not popular. However, recent work on semantics, pragmatics, and the psychology of emotions offers the tools for explaining speech that is expressive of emotions in general, and of hate in particular. I thus argue for the apparently trivial thesis that hate speech expresses hate.

Various authors contend that the opposite of this thesis is true. Jeremy Waldron considers talk of 'hate' a distraction, since the regulation of hate speech, which he favours, should not be the regulation of people's thoughts or feelings.[1] For Waldron, hate speech is harmful because it undermines people's assurance that they will not face violence, aggression, discrimination, or exclusion. Alexander Brown argues that it is a 'myth' that emotions, feelings, or attitudes of hate or hatred are part of the essential nature of hate speech.[2] His view is that the term 'hate speech' expresses instead a family-resemblance concept, and so we cannot identify one essential feature of hate speech.[3] Katharine Gelber offers an account of hate speech tailored to design policies to regulate 'sufficiently harmful speech' to warrant government regulation, although she says that the assumption that hate speech 'implies the presence of an emotion of hate' is mistaken.[4] These views illustrate two current lines of thought. On the one hand, there is widespread consensus against the idea that hate speech is essentially tied to an emotion of hate. On the other hand, there is no consensus about what hate speech is.

I think that there is such a thing as hate speech and that a theory of it can be developed. In this article, I do not give recommendations for policymaking. I do however consider that it is fundamental to understand what, if anything, is the underlying phenomenon. Understanding what hate speech is and how it works is necessary for appreciating when

it can reach the dangerous levels that would warrant legal regulation. The article primarily draws from current work in philosophy of language, with input from research about the psychology of emotions, to try to offer a characterization of the phenomenon of hate speech. A caveat: I will in general talk of contempt, disgust, anger, resentment, fear, etc. as emotions to avoid entering in discussions about whether these are emotions, standing attitudes, or sentiments. However, I will use 'sentiment' for the affective state expressed by hate speech, following the suggested usage in Fischer *et al.* in their 2018 article, 'Why We Hate'.[5]

   This article is organized as follows. In the next section, I briefly comment on some aspects of pragmatic accounts of harmful speech, namely Langton's[6] and McGowan's,[7] as speech that enacts norms that may have harmful effects. I point out some limitations in those accounts, and then I present a view of derogation recently defended by Marques and García-Carpintero,[8] which can be adapted to supplement those pragmatic theories. The main advantage of the resulting view is that the norms that hate speech enacts are provided by the normative appraisals constitutive of hate. What is needed for the application of this framework to hate speech is independent support for the idea that hate is a normative appraisal state. In the third section, I review some recent work on emotions. The positive claim of the article is that hate speech presupposes ongoing hate as a sentiment which 'organizes people's social world' against target groups and can motivate violent actions against them.[9] In the final section, I reply to some of Brown's objections that hate speech is not essentially expressive of hate.

2.   Pragmatic, Presuppositional, and Expressivist Accounts of Prejudicial Speech

The United Nations, in its International Convention on the Elimination of All Forms of Racial Discrimination, requires State Parties to condemn all propaganda and organizations 'which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to *justify or promote racial hatred and discrimination* in any form' (my emphasis).

   Other recent documents from the United Nations and other institutions warn of and denounce hate speech. In June 2019, Secretary General of the United Nations, António Gutterres, gave a speech where he announced the UN's plan for action on hate speech, saying:

> Over the past 75 years, hate speech has been a precursor to atrocity crimes, including genocide, from Rwanda to Bosnia to Cambodia.[10]

This supports the idea that any theory of hate speech must consider, at least, the cases of hate speech that preceded mass violence. Paradigmatic cases include the genocide in Rwanda that was preceded by a hate propaganda campaign against the Tutsis.[11] In the speech, Guterres clearly distinguished the causes of hate speech ('These root causes [of hate speech] include violence, marginalization, discrimination, poverty, exclusion, inequality, lack of basic education, and weak state institutions') and the effects of hate speech ('undermines social cohesion, erodes shared values, and can lay the foundation for violence, setting back the cause of peace, stability, sustainable development and the

fulfilment of human rights for all'), and he characterizes hate ideas and language as hate-fuelled:

> *Hateful and destructive views* are enabled and amplified exponentially through digital technology, often targeting women, minorities, and the most vulnerable. *Extremists* gather online and radicalize new recruits.

> In both liberal democracies and authoritarian regimes, some political leaders are bringing the *hate-fuelled ideas and language* of these [extremist] groups into the mainstream, *normalizing them, coarsening the public discourse and weakening the social fabric* [emphasis not in the original].[12]

The Council of the European Union published its resolution from 2008 on combatting certain forms and expressions of racism and xenophobia, where article 1 says that Member States of the EU will ensure that certain kinds of intentional conduct are punishable. The conduct that is punishable includes inciting to violence or hatred against a group by reference to race, colour, religion, descent, or national or ethnic origin. It also includes publicly condoning, denying, or grossly trivializing the crimes of genocide, crimes against humanity, and war crimes as defined in Articles 6, 7, and 8 of the Statute of the International Criminal Court, directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent, or national or ethnic origin when the conduct is carried out in a manner '*likely to incite to violence or hatred against such a group or a member of such a group*'.[13] What is at stake, then, is hate speech understood as the condoning or incitement of hatred or violence against certain groups. Hate speech, thus understood, may precede and lead to violence.

What an account of hate speech should explain then is this apparent datum: that a population that accepts and normalizes the linguistic derogation and dehumanization of a given group, under certain conditions, comes to normalize negative attitudes for that group and may come to condone harmful actions against its members.

In the remainder of this section, I present current pragmatic accounts of harmful speech and then introduce and expand on Marques and García-Carpintero's account of derogation to suggest how pragmatic accounts can be supplemented. According to us, an utterance that literally uses a slur presupposes contempt towards its target, thereby expressing contempt. Hate propaganda tends to involve the use of slurs and dehumanizing language, even if not all uses of slurs amount to hate speech.[14]

2.1. Pragmatic Accounts

In an influential paper,[15] Rae Langton suggested that Austin's distinction between an utterance's causal perlocutionary effects and the illocutionary act constituted by it can serve to argue that pornography subordinates women.[16] This suggestion was fruitful for the development of further work on harmful speech, for instance that of Mary Kate McGowan.[17] In her more recent work, which departs from her previous perspective on the topic, McGowan holds that ordinary bigoted speech enacts harmful norms by operating at a level other than the illocutionary – the level of 'parallel acts'.[18] The idea is that, parallel to whatever it does illocutionarily, any contribution to a norm-governed interaction enacts score changes, and therefore norms, for that specific interaction. A racist utterance made by an ordinary speaker on a crowded subway car (e.g. 'Go home. We don't need your kind here') illocutionarily

attempts to impose an obligation upon the addressee to 'go home' (i.e. it is an attempt to perform an illocutionary act of ordering, whose felicity depends, among other things, on whether the speaker has the relevant sort of authority) and, in parallel, would enact a number of score changes, and thus norms, for the ongoing interaction (e.g. making it okay to attack the targeted person). The speech would be constitutively harmful if the enacted norms prescribed harm and discrimination and those norms were followed.[19]

I do not have the space to discuss McGowan's view in detail here, but I will just say that I think it is not obvious that such speech can, if not questioned, succeed in enacting norms that make it okay to attack the targeted person. If I enter a crowded office space and shout, 'what does a person have to do to get some coffee around here?' and no one replies, I have not thereby made it okay for anyone to take my impertinence or my appetites seriously.[20]

More recently, Gelber has offered a 'conceptual understanding' of hate speech as systematically discriminatory, inspired by Langton's and McGowan's arguments. As she says,

> hate speech according to these arguments can be understood as a *discursive act* of discrimination, which operates on its targets in *constitutive* and *causal* ways to effect the denial of equal opportunities and rights.[21] [My emphasis]

Constitutively harmful speech would be distinguishable from speech that 'merely' has harmful effects: 'hate speech can harm by perpetuating covert rules which are capable of marginalizing targets and excluding them from public deliberation'.[22] Crucially, this systemic discrimination approach does not rely on the presence of an emotion of hate in the speaker. Gelber could be right about some of the systematic features of hate speech that require specific regulation: Is it directed at a group that is subjected to systemic discrimination? Does the speaker speak from a position of relative authority? Does it have wide support among the public?

I must, however, draw attention to a limitation of theories of hate speech based on existing systematic discrimination, in particular those that claim that hate speech targets groups identified by their protected characteristics. Hate speech can precede and correlate with hate crimes and mass violence. As Jonathan Maynard and Susan Benesch say, '[t]he forms of speech and ideology that catalyse mass violence, and the ways in which they do so, are strikingly similar across different cases'.[23] The similarities between cases transcend local systems of discrimination, and often the targets are not vulnerable groups identifiable by protected characteristics. Tutsis traditionally occupied government positions and had more 'power, social status, and influence' than Hutus for centuries under Tutsi monarchies and then under colonial rule.[24] If a set of protected categories were necessary to recognize hate speech, then cases like the RTML propaganda in Rwanda that contributed to the genocide would not count as such.[25] Since the Tutsi genocide in Rwanda was arguably motivated by hate propaganda, and it is a paradigmatic example mentioned in international plans for action against hate speech, any theory of hate speech that does not explain the Rwandan case has fatally failed in its aims. This is a problem for analyses based on pre-existing systems of discrimination.[26]

I think that there is a foundational problem with some of the assumptions of pragmatic accounts. Pragmatic views aim at giving a theory that explains how speech can acquire features that constitute it as harmful. However, Langton's explanation of illocutionarily harmful speech does not establish that it is harmful because of its illocutionary force. Yet, it is illocutionary force that constitutes what act is performed:

whether a request, a bet, or a promise. Langton cites the same UN document I mentioned earlier, and says, 'according to the UN description, racial hate speech *disseminates* ideas that are based on racial superiority; it *promotes* racial hatred and discrimination. It also *incites* …'.[27] The acts of disseminating, promoting, or inciting are not constitutively harmful or hateful. So, it is not what constitutes these speech acts as such that makes them harmful.

It is understandable that theorists would look for other sources of the constitutively harmful features of speech. The most obvious alternative is that it is the content that is promoted, incited, or disseminated that is harmful or discriminatory.[28] McGowan, for example, says that speech constitutes harm by causing it via the enactment of norms that prescribe that harm.[29] Constituting harm is a specific way of causing harm. It is thus only by looking to the effects of what is prescribed that we can retroactively look at the harm of the prescription. Inciting or promoting violence are wrongful acts because violence is harmful. But I think that accounts like McGowan's fail in characterizing such constitutively harmful features of speech, and this is the root of the problem for existing pragmatic accounts.

Part of the problem, I think, is that McGowan wants to cover acts and norms that are not direct threats, incitements to violence or aggression, insults, uses of slurs, or dehumanizing language. These are harmful because they are directives requiring[30] violence or threats of aggression. Nonetheless, McGowan wants to include speech that does not meet these conditions, but that can enact norms that may have discriminatory effects.[31] She illustrates the difference between 'merely causing' discrimination and enacting norms that cause discrimination by offering two examples of discrimination against red-haired people. One example involves causing discrimination by persuading others to act in discriminatory ways. The second example is meant to illustrate the harm caused by the enactment of a hiring policy:

> Suppose instead that I am an employer and implement a company hiring policy when I say, 'From now on, we no longer hire anyone with red hair'. This utterance will cause discriminatory conduct on the part of my employees, but it does so via the prescriptive force of the hiring policy enacted by my utterance.[32]

There are details about this story that are not laid out. We are invited to interpret it in a way that makes it true that the policy against hiring red-haired people is unfair differential treatment. But it need not be. And we cannot cast a particular policy implementation as constitutively harmful because, in certain conditions, it could have harmful effects. This makes it a bad explanation of constitutively harmful speech.

Suppose that the job in question is for a beach lifeguard on an island on the equator, where people prone to develop skin cancer should avoid sun exposure. The employer's decision to not hire people prone to get skin cancer in that line of work is not unfair, although it is differential treatment.[33] So there is nothing constitutively harmful in the speech act in question – it is an act that enacts a norm – and nothing about the content of the norm as such is harmful, discriminatory, or unfair. Whatever unfair features the enactment of a norm with the same content may have in another context are extrinsic to speech. Therefore, this pragmatic account does not explain why certain kinds of speech are constitutively harmful, if at all. If we want to explain how hate speech is constitutively harmful, we must find something intrinsic to it.

I will now introduce recent presuppositional accounts of derogatory discourse and suggest that they offer some of the resources needed to understand how speech can be constitutively harmful.

2.2.   *Expressive and Presuppositional Accounts of Derogatory Speech*

Truth-conditional theories of expressives, pejoratives, and slurs do not suffice to explain their meaning. As David Kaplan has argued, we need a further dimension of expressive meaning ('use-conditional', as opposed to truth-conditional) to characterize the meaning of words like 'damn', 'oops', or 'bastard', but also of slurs like 'honky'. We should not just ask for the truth-conditions of sentences containing these phrases but ask rather '[w]hat are the conditions under which the word is correctly and accurately used?'[34] After Kaplan, various theorists developed semantic accounts of expressives that capture the idea that a 'correct and accurate use' of an expressive signifies an emotion or affective state.[35]

There are various linguistic reasons to adopt an expressivist-presuppositional view about slurs and other epithets. First, slurs are derogatory, and this feature does not depend merely on the context of their use, or on speakers' intentions. Derogation is a part of slurs' meaning.[36] Moreover, slurs and other expressives display features that are typical of not-at-issue content.[37]

Robin Jeshion offers an expressivist use-conditional account and argues that pejoratives semantically express contempt, which she analyses as a moral emotion, after Michelle Mason.[38] Uses of a slur convey that its targets have lesser standing as humans, that they are undeserving of the full respect that is owed to persons *qua* persons, and they thereby dehumanize its targets.[39] Jeshion further argues that words can become derogatory by coming to conventionally mean the expression of contempt.[40]

Several authors persuasively argue that pejoratives and slurs presuppose their derogatory force (or meaning), based on evidence from occurrences of slurs embedded under negation, in conditionals, under modal operators, or in sentences with interrogative or directive mood. Philippe Schlenker,[41] for instance, held that a slur S presupposes a proposition of this form:

> *The agent of the context believes in the world of the context that the people in group targeted by S are despicable.*

This gives a clear-cut condition on a Stalnakerian context. For Stalnaker, to assert that φ is to propose that φ become common ground.[42] An accepted assertion comes to be presupposed afterwards and becomes common ground.[43] If Schlenker is right, an audience that understands and accepts an utterance with a slur updates its common ground with a presupposition of the form above.

However, Timothy Williamson has argued that presuppositions like the one Schlenker offers do not capture the normative status of slurs.[44] As Williamson said, for the case of 'Boche' and the presumed associated implicature associated with the word, 'The false implicature is that Germans are cruel, not that the speaker believes that they are cruel.'[45] This applies to Schlenker's proposed presupposition too. Take, for example, 'honky', a slur for white people in the United States.[46] According to Schlenker, if a speaker says (1), he asserts that the demonstrated person likes Fox News, and presupposes that he believes that white people are despicable:

(1)   That *Honky* Likes Fox News.

This leaves out something fundamental about the meaning of the slur. Hearers can accept that the demonstrated person likes Fox News, which is true. And hearers can accept that the speaker believes that white people are despicable, which presumably is also true. But hearers can find the slur strongly objectionable and reject (1). Hence, Schlenker's explanation does not capture what the audience rejects. Moreover, one of the explanatory advantages of a presuppositional account is lost – to explain how audiences that accept and normalize the derogation of a group through discourse may come to have contempt for the target group.

Josep Macià has a different presuppositional view that appears to avoid the above difficulty.[47] According to him, words like 'honky' presuppose that the speaker and audience are disposed to treat white people with contempt. If the audience is not bigoted, this presupposition is false, which would explain why the audience is not willing to accept (1). We can grasp the point with a comparison with the presupposition associated with a pronoun like 'she'. To accept an utterance like (2) is partly to accept that the referent of the context is female:

(2)   She's Not Coming on the Bike Ride.

Of course, it may be false that the demonstrated person is female, and the audience may refuse to accept (2) for this reason.

However, this solution also fails. To accept that the referent of the demonstrative pronoun is female is to accept all that there is to accept in (2), beyond accepting that she is not going on a bike ride. But to accept (1) is not merely to accept that we are disposed to treat the target of the slur with contempt. This point is easy to make. Suppose that the hearers have received diversity training, learned about structural racism, and now believe that they are racist. They believe the proposition that according to Macià is presupposed, and this is common ground. Hearers may nonetheless not condone or be disposed to treat anyone with contempt (otherwise the diversity training would have been pointless). It follows that to believe that we are disposed to treat others with contempt is not equivalent to regarding others with contempt. Thus, Macià's account also does not capture this core motivating feature of expressive-presuppositional views: that an audience who accepts the derogation of others through language may come to have contempt for the target group. The problem seems to be that we need to explain how slurs contribute to shared emotions, and not just to shared beliefs.

2.3.   *Context Update Beyond Shared Belief*

In recent years, new models of communication have acknowledged that discourse changes more than the set of propositions that are believed as true. It can also change questions that are under discussion and shared plans.[48] As an illustration of how such a model could work, consider the account of expressive-normative discourse of Alejandro Pérez Carballo and Paolo Santorio.[49] Their theory is inspired by Stalnaker's theory of assertion and by Gibbard's norm-expressivism.[50] Pérez Carballo and Santorio modify the Stalnakerian model in terms of sets of possible worlds by adding a complete system of norms *n* which would determine a three-way partition of possible courses of action – those that are

forbidden, permitted, or required according to *n*. Each assertion of a sentence is then assigned a set of world-norm pairs. The two following examples illustrate this. A set of norms is only relevant to the content of sentence (4):

(3)   Bob likes Fox News.
(4)   Fox News is despicable.
(5) $\{\langle w, n \rangle$: Bob likes Fox News in $w\}$.
(6) $\{\langle w, n \rangle$: Despising Fox News in w is permitted by $n\}$.

Common ground would now include both the possible worlds and the possible courses of actions that are left open by speakers' doxastic and normative acceptance attitudes.

Could such a model account for how derogatory discourse changes what audiences accept? The arguments in Section 2.2. show that slurs involve more than the mere acceptance of propositions as true. I think that accepting and condoning the use of slurs also involves more than accepting that there are possible courses of action that the interlocutors do not rule out. What is the difference? *Prima facie*, accepting a directive differs from accepting an expressive, and accepting a use of a slur, if Jeshion and others are correct, involves developing contempt for the targets of the slur. If slurs are expressive of contempt, then accepting the use of a slur, in general, can involve forming certain dispositions towards the slur's targets. Of course, contempt can be a motivator for certain discriminatory actions. But there is a difference between having contempt for some people and forming plans to socially exclude them. One can have the former without the latter, and vice versa.[51]

A currently prevailing view about emotions is that they are normative appraisal states.[52] As authors like Julien Deonna and Fabrice Teroni argue, emotions are intentional attitudes with their own psychological profiles and are appraisals that may be fitting, given their intentional objects.[53] For example, it is fitting to fear box jellyfish because they are extremely dangerous and deadly. Box jellyfish are the intentional object of fear, in this case. In contrast, it would not be fitting to fear, say, a daisy. Hence, this normative condition appraises fear of box jellyfish as fitting, and fear of daisies as unfitting.

Fear can be defined constitutively by its normative appraisal condition,[54] roughly expressed as *fear is the emotion that it is fitting to have towards dangerous things*. Being dangerous, thus, is the normative appraisal condition that something must meet for it to be fit to fear it. This condition is sometimes called the emotion's formal object. Additionally, many emotions involve certain action tendencies towards typical motivational goals: fear can lead to avoidance of its intentional object (although not necessarily so, or bravery would not be possible).[55] But how can emotions be shared?

Mikko Salmela and Michiru Nagatsu give a detailed account of collective emotions that links the intentional structure of joint actions and their underlying cognitive and affective mechanisms.[56] As they say, '[c]ollective emotions can function as both motivating and justifying reasons for jointly intentional actions, in some cases even without prior joint intentions of the participants. Moreover, they facilitate coordination in joint action.'[57] Emotions like resentment, contempt, disgust, fear, or anger can more easily spread among a community with which one identifies, through mechanisms like emotional contagion and empathy.[58]

What about contempt, then? The person that one feels contempt for is its intentional object. The formal object of contempt is the normative condition that makes it fit or

apt: being beneath equal consideration as a person.[59] In other words, it is fitting to feel contempt for someone only if they are beneath equal consideration as a person (it's another question if this is a condition anyone can meet). Contempt can motivate an indefinite number of possible action tendencies, e.g. the social exclusion or discrimination of the person in question. Of course, it need not motivate exclusionary or discriminatory actions. One can feel contempt for someone else but also have an overriding belief that no one should be socially excluded.

If slurs express contempt for their targets, then slurs cast those targets as unworthy of equal respect. This is the line that Marques and García-Carpintero follow.

> In order to accommodate a presuppositional view of pejoratives, we should hence add further illocutionary structure to contexts. The intentional objects of emotional states are the contents of this additional structure.[60]

Context updates, as we argue, will include the normative appraisal that identifies contempt for a target expressed through derogatory illocutionary acts. Derogation is defined with reference to contempt, schematically formulated thus:

> *Derogation*: one must derogate group G on account of their having features $F_1 \ldots F_n$ only if group G has $F_1 \ldots F_n$, and contempt fits a group with such features.[61]

To accept an utterance like (1), 'that honky is watching Fox News', is to accept the proposition that the demonstrated person is watching Fox News, but also to accept the fittingness of contempt towards the intentional object given by the meaning of the slur, i.e. white people. This is not the same as the explicit acceptance of the plan to treat group G with contempt.

Updating context with directives is thus different from updating context with expressives. Although the appraisals and norms that people's attitudes converge on may come to include plans of discriminatory actions towards the emotivational goals typical of contempt, that is contingent on further contextual circumstances, including whether people's commitment to rule out discrimination overrides the action tendencies that contempt typically generates. Accepting the expression of contempt is different from accepting that a series of possible courses of action are not to be ruled out.

This account of derogation supplements pragmatic accounts like Langton's and spells out what she stated as a programmatic suggestion:

> I want to propose, in an exploratory spirit, the idea that the phenomenon of accommodation might extend beyond belief – beyond conversational score, and common ground, as originally conceived – to include accommodation of other attitudes, including desire and hatred. My remarks here will inevitably be programmatic. But to convey the general idea: just as a hearer's belief can spring into being, after the speaker presupposes that belief, so too a hearer's desire can spring into being, after the speaker presupposes the hearer's desire; and so too a hearer's hatred can spring into being, after the speaker presupposes that hatred. Stalnaker's common ground can perhaps be extended to include not just common beliefs, and other belief-like attitudes, but common desires, and common feelings, as well. Speakers invite hearers not only to join in a shared belief world, but also a shared desire world, and a shared hate world.[62]

If speech can express emotions, then different kinds of expressive speech can express different attitudes, characterized by the specific appraisals that are constitutive of each. In the next section, I briefly draw from current work on hate to suggest that it can be combined with this framework of expressive discourse to better understand hate speech.

3.   Negative Emotions and the Hate Sentiment

In this section, I review some recent work on hate and related attitudes like contempt, disgust, anger, resentment, etc.

In a recent review of contemporary research on hate, Fischer *et al.* claim that hate is developed when others mistreat or humiliate someone, or when others' deliberate actions have become an obstruction to someone's goals, and that

> when individuals experience hate, they typically *perceive their hate target as having malicious intentions and being immoral, which is accompanied by feelings of lack of control or powerlessness*. Such *appraisals* are not the result of one specific action, but of a belief about the stable disposition of the hated person or group.[63] [My emphasis]

Citing research by Roseman, Wiest, and Swartz, the authors point out that an emotion's 'emotivational goal reflects what the emotion tries to bring about, and thus drives the emotional experience. Action tendencies are very closely associated with emotivational goals as they reflect the emotional impulse to act on a specific goal'.[64] They acknowledge that although hate is hard to characterize, it is distinguishable from closely related negative attitudes like anger or contempt not only by its specific constitutive normative appraisals (i.e. hate's formal object) but also by the emotivational goals and action tendencies that hate tends to generate:

> hatred is a hostile feeling directed toward another person or group that consists of malice, repugnance, and willingness to harm and even annihilate the object of hatred. Whereas anger implies a coercion goal, that is, the motive to change another person by attacking, confronting, or criticizing, contempt implies an exclusion goal (Fischer & Roseman, 2007), motivating individuals to exclude others from their social environment (Halperin, 2008; Halperin, Canetti, & Kimhi, 2012; Jasini & Fischer, 2018). Adopting a social functional perspective on emotions (e.g. Fischer & Manstead, 2016; Keltner & Haidt, 1999), we propose that the emotivational goal of hate is not merely to hurt, but to ultimately eliminate or destroy the target, either mentally (humiliating, treasuring feelings of revenge), socially (excluding, ignoring), or physically (killing, torturing), which may be accompanied by the goal to *let the wrongdoer suffer* (Ben-Ze'ev, 2008).[65]

Hate can be directed against a person or a group. Certainly, people can hate someone else, and groups of people can hate one person, for example, when that person is treated as a scapegoat. Hate can start as an acute hate event but evolve into a long-term enduring sentiment, for instance in intergroup contexts. Although not all emotions evolve into long-term sentiments,[66] Halperin *et al.* say that the hate sentiment in inter-group contexts is a stable and familiar 'hating' emotional disposition, and that 'a standing emotional

sentiment of hatred serves as the basis for unifying the in-group and offers a shared belief for overcoming conflicting views'.[67] The goal of the hate sentiment in intergroup contexts is to eliminate the outgroup from the ingroup's environment, for example, through an absolute separation from members of the other group. Recent theoretical and empirical work by Matsumoto *et al.* further suggests that anger, contempt, and disgust (the ANCODI hypothesis) are the basic elements of hatred and the key emotions associated with intergroup aggression.[68] The characteristics of hate sentiments in intergroup contexts are then, according to Fischer *et al.*:

(1) The negative appraisals of outgroup members as malevolent or malicious just by being members of that group,

(2) The action tendencies that go from revenge, social exclusion, or attacks to the destruction of the target group, and

(3) The motivation goals, such as the desire to harm, humiliate, or even kill the target.[69]

The first characteristic, the negative appraisals, is constitutive of what hate is. In other words, hate's formal object is being malevolent or malicious.

As Fischer *et al.* say, hate

organizes people's social world and helps strengthening the connection of the in-group ('in-group love') at the expense of various out-groups ('out-group hate'). To prevent future painful offenses by the hated group, the goal of the hate sentiment is to eliminate this group from their environment, for example through an absolute separation from members of the other group.[70]

This analysis is consistent with work by Mikko Salmela and Christian von Scheve, who argue that group-based emotions have contributed to the rise of populist right-wing parties in Europe.[71] They review research on economic and social changes in the last decades involving the transnational mobility of refugees from Africa and the Middle East, economic globalization and deregulation, and the privatization of social services. They then argue that psychological and attitudinal factors or collective mechanisms 'mediate between structural changes and the success of right-wing parties'. The attitudes at play are ressentiment (involving fear, insecurity, resentment, and hatred against perceived enemies) and emotional distancing from the social identities that evoke shame and in favour of positive social identities.

In the previous section, I claimed that if we are to explain what hate speech is constitutively, we should find something intrinsic to it. I also showed that some contemporary accounts explain how expressives can enact the appraisal norms that are constitutive of the affective state, emotion, or sentiment that is expressed. I am now able to suggest that we can make sense of the idea that hate speech is expressive of hate. We should be able to characterize paradigmatic cases of hate speech, like the hate propaganda campaign in Rwanda, while recognizing that not all hate propaganda leads to such mass violence. The group-based standing sentiment of hate, understood as intergroup hate, and experienced on behalf of one's own group while targeting an outgroup, appears to be the best candidate for the attitude expressed in the hate propaganda of Rwanda.[72] That is not to say that speakers cannot express their hate for someone on other grounds, or that propaganda in Rwanda did not involve additional speech acts.

Derogatory acts are constitutively harmful because they presuppose the fittingness of contempt against a group, i.e. they presuppose the aptness of the negative appraisal of members of a group as unworthy based on their (assumed) negative features. To recall, the negative appraisal condition, the formal object, of contempt is being beneath equal consideration as a person. This may generate a range of motivational goals and various possible exclusionary or discriminatory action tendencies, which would explain the potential of derogatory language to contribute to what Robin Jeshion called the 'bigotry formation' that follows the creation of new slurs.

If hate speech is expressive of intergroup hatred, then hate speech updates context with the presupposition of the fittingness of hate against an outgroup.[73] If accommodated, the presupposed appraisals are taken for granted and shared by interlocutors, especially among ingroup members that share social identities. The spread of hate speech would precede and correlate with hate crimes or mass violence (under propitious conditions) by presupposing hatred for a group as fitting. This explanation complements other pragmatic accounts that characterize harmful speech as speech that enacts harmful norms. It does so by identifying the intrinsic features of speech that make it pernicious.

By analogy with the case of derogatory illocutions, the expressive act of hate speech should be defined by the related constitutive appraisal norm for group hatred, which presupposes that hate is fitting towards members of the target group, based on traits that are assumed to define their group identity. Schematically, the norm would be

> *Hate speech*: we must hate members of group G on account of their group membership qua Gs, only if hatred fits such a group.

What makes hate speech pernicious, then, is the hatred that it normalizes and the action tendencies that it tends to generate. In the next section, I reply to Alexander Brown's objections to the idea that hate speech is essentially connected to hate.

4.   Reply to Brown's Objections

The framework offered here resists some objections to expressivism. I will focus on Brown's claim that hate speech does not essentially involve hate. He argues that it's a myth that attitudes of hate are essential to hate speech. More specifically, he argues against an analysis of the concept of hate speech in terms of three conditions: the speech itself or other expressive conduct, groups or classes of persons identified by protected characteristics, and the attitudes of hate. He then disambiguates four readings of 'hate speech' in terms of those three conditions and argues against each.

Brown points out, correctly, that under one understanding, people can say that hate speech is speech that is perceived as hateful by an audience.[74] The problem is that hearer reactions are not a good basis on which to regulate speech. Different people have varying sensibilities and regulating speech on their basis is inconsistent with the free speech doctrine. But the problem of identifying hate speech as that which is perceived as hateful is more serious than Brown acknowledges. There are many forms of speech that may be so perceived, but which are far from amounting to the kind of hate propaganda that the UN condemns.[75]

On the second possible reading, hate speech signifies or otherwise represents the speaker's emotions. As the reader can gather, this is the claim of this article, and so I'll

return to it later. The third possible reading understands hate speech as speech that is motivated by the hatred felt by the speaker for any members of groups or classes of persons identified by protected characteristics.[76] This is a bad alternative because it over-generates and would include speech that is in no way discriminatory, insulting, or threatening.

On the final reading, hate speech is speech that incites hatred toward any members of groups or classes of persons identified by protected characteristics.[77] Depending on how we understand it, we may either find straightforward objections, as I'll now consider, or find a view compatible with the expression of hate. On one understanding, this is 'speech that causes hate', and it is possible for thoughtful reasonable speech to cause an increase in hatred. For example, the announcement of electoral results may cause hatred in supporters of the losing party against supporters of the winning party, but this is not hate speech. If the losing party incited hatred, then the claim would have some weight. The International Covenant on Civil and Political Rights, article 20(2) says 'Any *advocacy* of national, racial or religious *hatred* that constitutes *incitement* to discrimination, hostility or violence shall be prohibited by law.' To incite, in this sense, does not merely cause harm; it mandates harmful conduct, possibly co-occurring with speech that is expressive of hatred.

Let me then reply to the objection that hate speech does not essentially involve attitudes of hate of a speaker for a group. This charge seems to assume that the description 'the emotions, feelings, or attitudes of hate of the speaker' is to be read as a *de re* definite description, such that, if the speaker does not feel hate, then the utterance is not expressive of hate.[78] But, as this article argues, this is based on general misunderstanding of expressive speech. Brown gives four apparent counterexamples, which I now discuss.

4.1. First Case: The Fundamentalist

A fundamentalist Christian says to women on a public street as they enter a lesbian bar, 'You lesbians are sinners in the eyes of God, you disobey the teachings of the Bible, and for this reason you will go to hell if you do not repent.' This utterance would not be 'an articulation or expression of emotions, feelings, or attitudes of hate or hatred' but rather '*an expression of deeply held religious beliefs or as some sort of religious exercise or rite of passage or even as an outpouring of feelings or sentiments of disgust or repulsion* that have been learned from parents or community leaders' (my emphasis).[79] Brown says that 'speech that carries the prejudiced (because contemptuous) message that homosexuals are morally inferior, or as a speech act that amounts to ranking, degrading, harassing or persecuting homosexuals' might qualify as hate speech.[80] But this is consistent with what the present account predicts and is not an objection.

There are two issues here. First, suppose that the above is an expression of religious belief. The statement can be a conventionalized way of implicating that lesbians are evil, impure, and deserving of hatred. Second, Brown concedes that there may be an 'outpouring of feelings or sentiments of *disgust* or *repulsion*'. This means that it is not obvious that he can deny that the fundamentalist is expressing hatred. By conceding that there are outpourings of disgust or repulsion, he is identifying what, according to Matsumoto *et al.*, is one of the components of hatred. He further identifies the most serious moral condemnation a Christian can make – lesbians are 'sinners in the eyes of God' – which is consistent with moral contempt ('because contemptuous', Brown says), again characteristics of the hate sentiment. I would add that taking the trouble to address random lesbian women on

the street further requires a good degree of anger. This case thus seems to illustrate the ANCODI hypothesis.

If the appraisals that hate speech against lesbians makes are accommodated, then lesbians come to be viewed as evil – 'You are sinners, and you'll go to hell!' Moreover, under certain conditions, some actions can become permissible (or required): revenge, social exclusion, or attacks to achieve emotivational goals like the desire to harm or humiliate. Lesbians who have had these experiences will recognize both the speech and the attacks.

4.2.   Second Case: Scientist

Imagine a scientist who publishes a controversial paper with the results of his research about the relative intelligence of African Americans and which includes the sentence, 'African Americans tend to have lower IQs than white Americans.' As Brown says, we could think that this is hate speech 'by virtue of the fact that in a very public way it rehearses and supports … a particular negative stereotype or social stigma about African Americans'.[81] This case falls under what Mari Mikkola calls 'hard cases'.[82] She follows David Brink who argues that a case of this kind may count as discriminatory speech, in that it 'reflects group stereotypes and represents groups or their members as inferior by virtue of these stereotypes'.[83] Whether such discriminatory speech is additionally expressive of hatred may depend on additional circumstances, perhaps involving similar reasoning as that given about the Christian fundamentalist. It is nonetheless useful to have separate categories.

4.3.   Third Case: Disablists

Some disablists could use the insults 'retard', 'cripple', and 'freak' to express their contempt or disdain for people with physical or mental disabilities, revealing that they look down on them, believing them to be worthless or beneath consideration, without at the same time 'articulating, symbolizing or representing emotions, feelings, or attitudes of hate or hatred toward such people'.[84] As in previous cases, the expression of contempt and disdain, together with the belief that the people in question are worthless or beneath consideration, may count as expressions of hate and again are compatible with my account. But these may be just derogatory acts that do not amount to hate speech, for instance if they are not incitements to hatred and violence.

4.4.   Fourth Case: Sexist Husbands

A group of married men can talk together about their wives and women in general. 'Together they espouse a range of derogatory ideas and negative stereotypes about women (e.g. "a woman's place is in the home", "the man is in charge", "women are no good at practical things around the house"), and at the same time they neither question these ideas nor offset them with more positive stereotypes. But just suppose for the sake of argument that the men do not in fact harbour any emotions, feelings, or attitudes of hate or hatred toward their wives.'[85] It is not clear what this case is supposed to show. First, it is not clear that any sexist or bigoted conversation should amount to hate speech, and it is not clear that this should be something about which speakers 'instinctively' have views. Second,

presuppositional accounts of derogation account for this kind of speech without needing to categorize it as hate speech.

5.   Conclusion

I have argued that hate speech is speech that is constitutively prejudicial because it is expressive of hatred. I have offered a view of expressively normative speech that modifies Marques and García-Carpintero's account of derogation, and I proposed that speech that is expressive of an emotion or sentiment presupposes that emotion as fitting towards its intentional object. If the presupposition is accommodated, it modifies context with the normative appraisals that constitute the expressed attitude. Under propitious circumstances, shared goals and actions characteristic of that attitude can also become common ground. Current research on intergroup hate suggests that hate is indeed an appraisal state, and that intergroup hate can contribute to structure people's social world. This supplements other pragmatic accounts that do not always make clear why prejudicial speech is constitutively harmful. Finally, I have argued that objections to the idea that hate speech is essentially expressive of hatred tend to misunderstand the nature of expressive discourse and the nature of the affective states expressed.[86]

Teresa Marques, Department of Philosophy, Universitat de Barcelona, LOGOS Research Group, and BIAP – Barcelona Institute of Analytic Philosophy, Barcelona, Spain.
teresamatosferreira@ub.edu

NOTES

1 Waldron, Jeremy. 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press, pp. 34–6.

2 Brown, Alexander. 2017a. "What is Hate Speech? Part 1: The Myth of Hate." *Law and Philosophy* 36(4): 419–68.

3 Brown, Alexander. 2017b. "What is Hate Speech? Part 2: Family Resemblances." *Law and Philosophy* 36(5): 561–613.

4 Gelber, Katharine. 2021. "Differentiating Hate Speech: A Systemic Discrimination Approach." *Critical Review of International Social and Political Philosophy* 24(4): 393–414.

5 Fischer, Agneta, Eran Halperin, Daphna Canetti, and Alba Jasini, 2018. "Why We Hate." *Emotion Review* 10(4): 309–20.

6 Langton, Rae. 2012. "Beyond Belief: Pragmatics in Hate Speech and Pornography." In *Speech and Harm: Controversies Over Free Speech*, edited by M. K. McGowan and I. Maitra, 72–93. Oxford: Oxford University Press.

7 McGowan, Mary Kate. 2009. "Oppressive Speech." *Australasian Journal of Philosophy* 87(3): 389–407.

8 Marques, Teresa, and Manuel García-Carpintero. 2020. "Really Expressive Presuppositions and How To Block Them." *Grazer Philosophische Studien* 97(1): 138–58.

9 Fischer *et al.* op. cit.

10 Guterres, António. 2019. "Secretary-General's Remarks at the Launch of the United Nations Strategy and Plan of Action on Hate Speech." https://www.un.org/sg/en/content/sg/statement/2019-06-18/secretary-generals-remarks-the-launch-of-the-united-nations-strategy-and-plan-of-action-hate-speech-delivered. Accessed on January 2022.

11 The correlation claim between hate propaganda and mass violence in Rwanda is established in Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *Quarterly Journal of Economics* 129(4): 1947–94.

12 Guterres op. cit.

13 "Council Framework Decision 2008/913/JHA of 28 November 2008 on Combating Certain Forms and Expressions of Racism and Xenophobia by Means of Criminal Law." *Official Journal of the European Union*, L 328/55. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32008F0913. Accessed on January 2022.

14 Authors working on hate and dangerous speech highlight the role of derogatory and dehumanizing language in it, for instance Maynard, Jonathan, and Susan Benesch. 2016. "Dangerous Speech and Ideology: An Integrated Model for Monitoring and Prevention." *Genocide Studies and Prevention: An International Journal* 9 (3): 70–95, p. 80, or Tirrell, Lynne. 2012. "Genocidal Language Games." In *Speech and Harm: Controversies Over Free Speech*, edited by Ishani Maitra and Mary Kate McGowan, 174–221. Oxford University Press, especially section 3. Authors writing on slurs also emphasize how slurs contribute to dehumanize and discriminate vulnerable groups, e.g. Robin Jeshion says that 'slurs function to derogate or dehumanize, by which I mean, that they function to signal that their targets are unworthy of equal standing or full respect as persons, that they are inferior as persons.' Jeshion, Robin. 2013. "Expressivism and the Offensiveness of Slurs." *Philosophical Perspectives* 27(1): 231–59. However, although theorists like Maynard and Benesch, or Tirrell, consider that derogatory speech is a component of dangerous speech or genocidal speech (which are extreme forms of hate speech), they do not say that all uses of derogatory language amount to dangerous or genocidal language. I agree that not all derogation reaches the level of hate speech. Antonio Guterres's address to the UN characterized hate ideas and language as hate fuelled, and other official documents characterize hate speech as that which incites to violence and hatred. I take it that it is obvious that not all uses of epithets, pejoratives, or slurs display the features that the UN or the EU condemn in hate speech: hate-fuelled incitements to violence or hatred against social groups.

15 Langton, Rae. 1993. "Speech Acts and Unspeakable Acts." *Philosophy and Public Affairs* 22: 293–330.

16 Austin, J.L. 1962. *How to Do Things with Words*. Oxford: Oxford University Press.

17 McGowan 2009 op. cit.

18 McGowan, Mary Kate. 2019. *Just Words: On Speech and Hidden Harm*. Oxford: Oxford University Press, p. 94.

19 I am grateful to an anonymous reviewer for pressing me on this point.

20 An anonymous reviewer pointed out that the speaker's words do not presuppose in any way 'it is okay for you to take my impertinence/appetites seriously'. Rather, in saying, 'What does a person have to do to get some coffee around here?' the speaker seems to presuppose that impertinent behaviour is generally okay (or okay for them) in that office space. Provided that nobody objects (if McGowan is right), their words would enact a norm to that effect. I agree that the words do not carry that presupposition, semantically. And I also agree that it is possible that, on specific occasions, audiences can come to accommodate something like the presupposition that insolent behaviour is okay. But I think that a lack of response from the audience does not thereby enact the norm that it is permissible for me to be impertinent. If all my colleagues ignore me and no one shows me where the coffee is, I've been politely put in my place. Indeed, this is often an effective parenting strategy to deal with toddlers. And as some say about engaging with controversies on Twitter – 'do not feed the trolls', i.e. do not reply.

21 Gelber op. cit., p. 6.

22 Gelber op. cit., p. 15.

23 Maynard and Benesch theorize dangerous speech as speech that is capable of encouraging approval of violence by an audience. I think we can understand their project as a study of the conditions in which hate speech correlates with hate crimes or mass violence. It is only in specific circumstances, as they show, that such speech becomes dangerous. Maynard and Benesch op. cit., p. 71.

24 Tirrell op. cit., p. 178.

25 Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Theory and Evidence from the Rwandan Genocide." *Quarterly Journal of Economics* 129(4): 1947–94.

26 See for instance Kindermann, Dirk. 2021. "Against 'Hate Speech'." Vienna: University of Vienna. Unpublished manuscript.

27 Langton op. cit., p. 75.

28 Mikkola, Mari. 2021. "Discriminatory vs Hate Speech: Wherein Lies the Difference?" Unpublished manuscript, argues that the difference between hate speech and other types of prejudicial speech is content based. Amsterdam: University of Amsterdam.

29 McGowan 2019 op. cit., p. 20.

30 I use 'requires' for lack of a better term to phrase the strength of the directives in incitements to violence. An anonymous reviewer suggested 'gives the audience non-obligatory reasons', but I do not think this is the right

way to characterize incitements to violence, because 'gives reasons' intellectualizes the contents of such acts. Moreover, 'non-obligatory' seems to leave too much open to rational deliberation, which seems uncharacteristic of the power of incitements to violence – just imagine a mob shouting 'Kill!' during a public execution.

31 In contrast, I believe that for instance derogatory acts are harmful because they presuppose the fittingness of contempt towards people and therefore imply that people are *unworthy of equal consideration as persons*. I return to this below.

32 McGowan 2019. op. cit., p. 24.

33 See Lippert-Rasmussen, Kasper. 2014. *Born Free and Equal? A Philosophical Inquiry into the Nature of Discrimination*. Oxford: Oxford University Press.

34 Kaplan, David. 1999. "The Meaning of 'Ouch' and 'Oops'." Los Angeles, CA: University of California Los Angeles. Unpublished manuscript, p. 5.

35 See, for example, Potts, Chris. 2007. "The Expressive Dimension." *Theoretical Linguistics* 33(2): 165–98.

36 See Cepollaro, Bianca. 2015. "In Defense of a Presuppositional Account of Slurs." *Language Sciences* 52: 36–45. See also Díaz-Legaspe, Justina. 2020. "What is a Slur?' *Philosophical Studies* 177(5): 1399–422. Díaz-Legaspe argues that slurs are words that are grammatically marked as strongly derogatory.

37 On the projective behaviour of slurs, see Tonhauser, Judith, David Beaver, Craige Roberts, and Mandy Simons, 2013. "Toward a Taxonomy of Projective Content." *Language* 89: 66–109; Camp, Elizabeth. 2018. "A Dual Act Analysis of Slurs." In *Bad Words: Philosophical Perspectives on Slurs*, edited by David Sosa, 29–59. Oxford: Oxford University Press; Sennet, Adam, and David Copp, 2015. "What Kind of a Mistake Is It to Use a Slur?" *Philosophical Studies* 172: 1079–104. On conventional implicature accounts, see Potts op. cit., and Potts, Christopher. 2012. "Conventional Implicature and Expressive Content." In *Semantics: An International Handbook of Natural Language Meaning*, edited by Claudia Maienborn, Klaus von Heusinger, and Paul Portner, 2516–35. Berlin: Mouton de Gruyter. Other presuppositional accounts are mentioned below.

38 Jeshion op. cit.; Mason, Michelle. 2003. "Contempt as a Moral Attitude." *Ethics* 113: 234–72.

39 Jeshion op. cit., p. 131.

40 Jeshion, Robin. 2016. "Slur Creation, Bigotry Formation." *Phenomenology and Mind* 11: 130–39.

41 Schlenker, Philippe. 2007. "Expressive Presuppositions." *Theoretical Linguistics* 33: 237–45.

42 Stalnaker, Robert. 2002. "Common Ground." *Linguistics and Philosophy* 25: 701–21.

43 Stalnaker, Robert. 1978. "Assertion." In *Syntax and Semantics*, vol. 9, edited by P. Cole, 315–32. New York: Academic Press.

44 Williamson, Timothy. 2019. "Reference, Inference, and the Semantics of Pejoratives." In *The Philosophy of David Kaplan*. edited by J. Almog and P. Leonardi, 137–58. Oxford: Oxford University Press.

45 Williamson op. cit., p. 152.

46 According to Wikipedia, it could have evolved from 'hunkie', a variation on 'bohunk', a derogatory term for Slavic and Hungarian immigrants (https://en.wikipedia.org/wiki/Honky, accessed on November 20, 2021).

47 Macià, Josep. 2014. "Expressive Meaning and Presupposition." Handout for a talk at the *Names, Demonstratives, and Expressives* conference in Gargnano, Italy.

48 See for instance, Roberts, Craige. 2012. "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics." *Semantics & Pragmatics* 49(5): 1–69, and Roberts, Craige. 2018. "Speech Acts in Discourse Contexts." In *New Work on Speech Acts*, edited by D. Fogal, D.W. Harris, and Matt Moss, 317–59. Oxford: Oxford University Press; Portner, Paul. 2016. "Imperatives." In *Cambridge Handbook of Semantics*, edited by M. Aloni and P. Dekker, 593–627. Cambridge: Cambridge University Press; García-Carpintero, Manuel. 2015. "Contexts as Shared Commitments." *Frontiers in Psychology* 6: 1932; García-Carpintero, Manuel. 2010. "On the Nature of Presupposition: A Normative Speech Act Account." *Erkenntnis* 85(2): 269–93.

49 Pérez Carballo, Alejandro, and Paolo Santorio, 2016. "Communication for Expressivists." *Ethics* 126(3): 607–35. For discussion, see also Marques, Teresa. 2021. "Illocutionary Force and Attitude Mode In Normative Disputes." *Metaphilosophy* 52(3–4): 449–65.

50 Gibbard, Alan. 1990. *Wise Choices, Apt Feelings*. Oxford: Oxford University Press.

51 To illustrate the difference, I recommend viewing the two-minute hate scene of the movie adaptation of Orwell's *1984*, where the crowd shouts hateful epithets. Although the character named Julia shouts 'death', she is not putting forward any plan of action. The clip is available on YouTube: https://youtu.be/XvGmOZ5T6_Y

52 Versions of this view about the nature of emotions have been defended by authors like Mulligan, Kevin. 1998. "From Appropriate Emotions to Values." *The Monist* 81(1): 161–88; D'Arms, Justin, and Daniel Jacobson, 2000. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61(1): 65–90; Deonna, Julien, and Fabrice Teroni, 2012. *The Emotions*. London: Routledge. An

anonymous reviewer asks whether accounts of emotions as appraisal states conflict with Gibbard-style meta-ethical expressivism. Authors like Gibbard seek to characterize moral judgments in terms that do away with the acceptance of propositions that refer to moral properties. Theories of emotions as appraisal states claim that an emotion like contempt is appropriate or apt just in case it is directed at someone unworthy of equal respect. But stating the aptness conditions for contempt reintroduces normative or evaluative properties, which the meta-ethical expressivist wished to do away with. Now, that may be a problem for the metaethical expressivist, but it is not in itself a reason against treating emotions as appraisal states. In any case, and as D'Arms and Jacobson op. cit. point out, we should take care not to commit a moralistic fallacy when talking of the appropriate conditions of emotions. One can have an emotion like contempt, therefore appraising its object as inferior or unworthy, while not judging that its object is unworthy. The existence of expressives in natural languages – phrases that express affects or emotions as appraisal states – stands independently from metaethical expressivism.

53 Deonna, Julien, and Fabrice Teroni, 2015. "Emotions as Attitudes." *Dialectica* 69: 293–311.

54 Several authors have rule-based accounts of speech acts, like assertion. A constitutive norm discriminates between a speech act and all other speech acts. The norm also indicates the conditions under which the speech act is correct. (Note that here what is at stake is correctness according to its constitutive norm, and not felicity. An assertion may be correct *qua* assertion, while infelicitous, perhaps because it is irrelevant.) Williamson argues that flat-out assertions are constitutively defined by a knowledge rule (Williamson, Timothy. 1996. "Knowing and Asserting." *Philosophical Review* 105: 489–523). *Mutatis mutandis*, the constitutive appraisal state of a conative attitude discriminates that attitude from all other possible attitudes; moreover, it identifies the conditions under which the attitude is apt or appropriate.

55 I am grateful to Filippo Contesi for insisting that action tendencies need not lead to actual actions. Salmela and Nagatsu further argue that in the case of collective emotions the action tendencies include dispositions for shared action (Salmela and Nagatsu op. cit., p. 36).

56 Salmela, Mikko, and Michiru Nagatsu, 2016. "Collective Emotions and Joint Action." *Journal of Social Ontology* 2(1): 33–57.

57 Salmela and Nagatsu op. cit., p. 33. See also Salmela, Mikko. 2012. "Shared Emotions." *Philosophical Explorations* 15(1): 33–46.

58 Salmela and Nagatsu op. cit., p. 40.

59 See Teroni, Fabrice. 2007. "Emotions and Formal Objects." *Dialectica* 61(3): 395–415.

60 Marques and García-Carpintero op. cit., p. 146.

61 Marques and García-Carpintero op. cit., p. 147.

62 Langton op. cit., p. 86.

63 Fischer *et al.* op. cit., p. 310.

64 Fischer *et al.* op. cit., p. 311.

65 Fischer *et al.* op. cit., p. 311.

66 Ben-Ze'ev, Aaron. 2018. "Is Hate Worst When It Is Fresh? The Development of Hate over Time." *Emotion Review* 10(4): 322–24.

67 As Halperin *et al.* say, in some situations there is a connection between hate and its various active manifestations: outgroup exclusion, motivation to fight and kill, etc. Halperin, Eran, Daphna Canetti, and Shaul Kimhi, 2012. "In Love with Hatred: Rethinking the Role Hatred Plays in Political Behavior." *Journal of Applied Social Psychology* 42(9): 2231–56, p. 2233.

68 Matsumoto, David, Hyisung C. Hwang, and Mark G. Frank, 2017. "Emotion and Aggressive Intergroup Cognitions: The ANCODI Hypothesis." *Aggressive Behavior* 9999: 1–15. Contrast with Fischer *et al.* op. cit., p. 312.

69 After Fischer *et al.* op. cit., p. 311.

70 Fischer *et al.* op. cit., p. 311.

71 Salmela, Mikko, and Christian von Scheve, 2017. "The Emotional Roots of Right-Wing Political Populism." *Social Science Information* 56(4): 567–95.

72 For work on group-based hatred and the role of hatred in political behaviour, see Halperin, Eran. 2008. "Group-Based Hatred in Intractable Conflict in Israel." *Journal of Conflict Resolution* 52(5): 713–36; Fischer *et al.* op. cit., p. 313; Haq, Hina, Saad Shaheed, and Achim Stephan, 2020. "Radicalization Through the Lens of Situated Affectivity." *Frontiers in Psychology* 1(l): 205; or Matsumoto, David, Mark G. Frank, and Hyisung C. Hwang. 2015. "The Role of Intergroup Emotions in Political Violence." *Current Directions in Psychological Science* 24(5): 369–73.

73 Deonna and Teroni 2012 op. cit., especially Chap. 7.

74 Brown 2017a op. cit., p. 447.

75 As anecdotal evidence, consider that the Portuguese right-wing dictator António de Oliveira Salazar made a speech in the 1930s justifying censorship to protect the 'souls tormented by doubt and the negativism of the century'. Nogueira, Franco. 1977. *Salazar: Estudo biográfico*. Atlântida Editora, p. 268.

76 Brown 2017a op. cit., p. 455.

77 Brown 2017a op. cit., p. 458.

78 Brown 2017a op. cit. Refers to the attitudes of the speaker, for instance, on p. 440.

79 Brown 2017a op. cit., p. 450.

80 Brown 2017a op. cit., p. 452.

81 Brown 2017a op. cit., p. 452.

82 Mikkola op. cit.

83 Brink, David. 2001. "Millian Principles, Freedom of Expression, and Hate Speech." *Legal Theory* 7: 119–57.

84 Brown 2017a op. cit., p. 452.

85 Brown 2017a op. cit., p. 452.