Croatian Journal of Philosophy Vol. XXVI · No. 77 · 2026

Institute of Philosophy

ISSN 1333-1108

Croatian Journal of Philosophy

Vol. XXVI · No. 77 · 2026

Croatian Journal of Philosophy

1333-1108 (Print)
1847-6139 (Online)

Published by
Institute of Philosophy
Ulica grada Vukovara 54/IV, 10000 Zagreb, Croatia
www.ifzg.hr

Available online at https://cjp.ifzg.hr
and https://hrcak.srce.hr/en/cjp

Croatian Journal of Philosophy is published three times a year. It publishes original scientific papers in the field of philosophy.

Croatian Journal of Philosophy is indexed in *The Philosopher's Index, PhilPapers, Scopus, ERIH PLUS* and in *Arts & Humanities Citation Index* (*Web of Science*).

Croatian Journal of Philosophy is published with the support of the Ministry of Science, Education and Youth of the Republic of Croatia.

Instructions for Contributors

All submissions should be made through our online submission system: https://ojs.srce.hr/index.php/cjp/about/submissions. Submissions must be in English and formatted to be double-spaced with suitably wide margins, an A4 page size, and automatic page numbering.

Articles are normally no longer than 8,000 words of main text including bibliography. The Journal will consider longer papers, but, once these limits are exceeded, authors should bear in mind the editorial policy that the acceptance bar raises with increasing length.

Manuscripts should be compiled in the following order: cover page; title; abstract (not exceeding 200 words); keywords (3 to 6); main text; appendices (as appropriate); references.

All the authors of a paper should include their full names, affiliations, postal addresses, telephone and fax numbers and email addresses on the cover page of the manuscript. If a paper is co-written, one author should be identified as the Corresponding Author. The cover page must be submitted as a separate document. All submitted manuscripts must be prepared for blind review, with revealing acknowledgements and self-identifying references removed.

Sources are cited in the text by the author's last name, the publication date of the work cited, and a page number if needed, e.g. (Barber 2007: 324). Full details appear in the reference list in which the year of publication appears immediately after the author's name:

Barber, A. 2007. "Linguistic Structure and the Brain." *Croatian Journal of Philosophy* 21 (7): 317–341.

Williamson, T. 2013. *Identity and Discrimination*. Oxford: Wiley-Blackwell.

The publication of a manuscript in the Croatian Journal of Philosophy is expected to follow standards of ethical behavior for all parties involved in the publishing process: authors, editors, and reviewers. The journal follows the principles of the Committee on Publication Ethics (https://publicationethics.org/resources/flowcharts).

CROATIAN JOURNAL OF PHILOSOPHY

Vol. XXVI · No. 77 · 2026

Introduction

This issue brings together six articles that, from different philosophical perspectives, examine questions concerning language, interpretation, political institutions, privacy, personal identity, and moral respect. Although thematically diverse, the contributions are connected by a shared concern with the ways in which we understand persons, social practices, and the normative relations in which we participate.

The issue opens with Y. Sandy Berkovski's "Dehumanisation Through Slurs", which addresses the claim that slurs dehumanise their targets. Berkovski considers several versions of this claim, beginning with the idea that dehumanisation is a form of disrespect and that the use of a slur therefore amounts to treating, or thinking of, its target with disrespect. He then turns to the suggestion that certain slurs are connected, in thought and action, with dehumanising treatment in cases of genocide. Berkovski argues that both the theoretical justification and the empirical support for these views are problematic, and concludes that the prospects for aligning slurs and dehumanisation are limited.

Carola Barbero and Fabrizio Calzavarini, in "Repeatable Experiences of Music and Texts", revisit Peter Kivy's analogy between reading literary texts and musical performance. While Kivy argues that reading, unlike musical experience, is not intrinsically repeatable because the discovery of plot loses its force after the first reading, Barbero and Calzavarini challenge this view. Drawing on theoretical considerations and empirical evidence, they argue that literature, like music, has traditions and structures of repetition, and that rereading can deepen appreciation and understanding rather than merely repeat an exhausted experience.

The third article, Enes Kulenović's "Athenian Democracy: A Debate between its Advocates and Critics", reconstructs and critically evaluates ancient debates concerning democratic rule in Athens. Focusing on questions of efficiency, stability, and justice, Kulenović examines the arguments of both critics and defenders of democracy. Since the critical arguments are generally more extensive and philosophically developed, the article reconstructs pro-democratic positions from the writings of democracy's critics as well as from historical scholarship, offering a balanced assessment of the strengths and limitations of both sides.

In "In Defence of the Right to Out Others", Miloš Kovačević addresses the ethics of outing in the context of privacy, sexual orientation, stigma, and collective action. The article starts from the observation that, unlike many visible stigmatized characteristics, sexual orientation can often be concealed, allowing individuals to reduce personal costs while con-

tributing to a broader collective action problem. Through a discussion of Jean Louise Cohen's reflexive conception of privacy and Judith Jarvis Thomson's reductionist account, Kovačević argues against understanding privacy as a right that generally prevents others from legally obtaining and disseminating information about someone's sexual orientation when that information is not subject to confidentiality.

Samuel Kahn's "The Transgender Experience and the Limits of Physiological Continuity Theories" brings together the philosophy of personal identity and the philosophy of transgender identity. Kahn examines a tension between physiological continuity theories of personal identity and one influential way of framing the transgender experience. After outlining both the relevant commitments of physiological continuity theories and the thesis about transgender experience that creates the difficulty, he considers two possible ways forward. Although each faces challenges, Kahn argues that either can avoid the apparent conflict.

The issue closes with Qing Zhao's "Self-Respect Is Incomplete Without Respect for Others", which develops an account of self-respect within Kantian ethics. Zhao argues that self-respect, understood as a duty to oneself grounded in the dignity of moral agency, cannot be fully understood in isolation from one's relations to others. By examining vices such as arrogance, defamation, and ridicule, and by developing an intersubjective account of dignity, the article shows that respect for others is not merely compatible with self-respect but constitutive of it.

TVRTKO JOLIĆ
Institute of Philosophy, Zagreb, Croatia

Croatian Journal of Philosophy
Vol. XXVI, No. 77, 2026
https://doi.org/10.52685/cjp.26.77.1
Received: September 15, 2025
Accepted: March 27, 2026

Dehumanisation Through Slurs

Y. SANDY BERKOVSKI
Bilkent University, Ankara, Turkey

A claim sometimes made in the literature on slurs is that slurs dehumanise their targets. Here, I look at several different versions of this claim. I begin by examining and ultimately rejecting the idea that dehumanisation is a form of disrespect, and that to use a slur is to treat the target, or to think of her, with disrespect, hence, to dehumanise her. Then I look at the idea that some select slurs are aligned, in thought and in action, with dehumanising treatment in genocide. I argue that the theoretical justification of this claim is confusing, while the empirical evidence is sketchy at best. I conclude by arguing that, in general, the prospects for aligning slurs and dehumanisation are dim.

Keywords: Dehumanisation; slurs; respect; genocide; perception.

I

A claim sometimes made in the literature on slurs is that slurs dehumanise their targets. Here, I look at several versions of this claim that were defended at any significant length. In §II, I examine and ultimately reject the idea that dehumanisation is a form of disrespect, and that to use a slur is to treat the target, or to think of her, with disrespect, hence, to dehumanise her. In §III, I examine the idea that some select slurs are aligned, in thought and in action, with demonstrably dehumanising treatment in genocide. I argue that the theoretical justification of this claim is confusing, while the empirical evidence is sketchy at best. I conclude in §IV by arguing that, in general, the prospects for aligning slurs and dehumanisation are dim.

II

In this section, I address Robin Jeshion's attempt to link dehumanisation to a denial of respect. I argue that it is unsuccessful, whether or not we think of respect along broadly Kantian lines.

To explain how slurs dehumanise, Jeshion distinguishes between dehumanising thought and dehumanising action, and then also between strong and weak senses of dehumanisation. Paraphrasing only slightly, here are the four notions:[1]

(1) a. x thinks of y in a dehumanising way in the *weak* sense iff x thinks of y as a human, but an inferior one, and unworthy of full respect.
b. x treats y in a dehumanising way in the *weak* sense iff x treats y as an inferior human unworthy of full respect.
c. x thinks of y in a dehumanising way in the *strong* sense iff x thinks of y as someone who is not fully human (less than human), and who completely lacks the moral status of a person.
d. x treats y in a dehumanising way in the *strong* sense iff x treats y as someone who is not fully human, and who completely lacks the moral status of a person.

Jeshion defends the weak versions (1a) and (1b). To call individuals with a slur is to think of them, and treat them, as inferior humans unworthy of full respect (Jeshion 2018: 103). Another component of her idea of dehumanisation is the characteristic emotion of contempt. Users of slurs express contempt toward their targets, and, in doing so, they regard them as "low in worth" (Jeshion 2018: 92, 94).

There are several reasons why this is a dubious view. Jeshion says that to be an "inferior human" is to have a "lower status along the dimension of *personhood* or *humanity in the moral sense*" (Jeshion 2018: 80, her italics). Now, suppose we learn that A was a ruthless murderer on a major scale who displayed few moral virtues in his public or private life. And suppose that B is a leading scientist, a committed parent, and a humanitarian. Aren't we entitled to think of B as a morally superior human? We are indeed. We can rank people according to their moral qualities, just as we rank them according to their other qualities. If the notion of humanity is associated with certain behaviours and capacities, and if manifesting these behaviours and possessing these capacities are scalable, then being human in the moral sense is scalable too. The same goes for Jeshion's own example (derived from Michelle Mason's work) of a woman who comes to despise her husband (Jeshion 2018: 91). The woman now thinks less of her husband as a person. Compared to other humans, he is inferior: it turns out that he has several bad moral qualities.[2]

[1] See Jeshion (2018: 79). Although Jeshion repeatedly refers to David Livingstone Smith's work on dehumanisation, his conception of it is based on essentialism, rather than denial of respect: see Smith (2011: 95ff) and also Smith (2020: 63ff). A view more in line with Jeshion's explications is in Mikkola (2016: 164). That dehumanisation based on stereotypes (and possibly some form of essentialism) comes in degrees is argued, in a different setting, in Fiske (2021), Fiske (2011: 1–27). I return to Smith's and Fiske's ideas in §IV.

[2] To clarify, we do not here take a stance on how the putative ranking should be determined. We are rather saying that, however you think of moral value, it is

This sort of idiom is plausible, but is it at all related to dehumanisation? When we despise someone, do we thereby "dehumanise" him? I think the clear answer is no. That is not how we think or speak. By the same token, we could regard the murderer as a morally inferior human being. The woman could regard her husband as a morally inferior being. These facts aren't sufficient for us to conclude that the individuals in question are being dehumanised. Therefore, while we may agree that to use a slur toward *x* is to despise *x* and to regard him as having low worth, these attitudes won't amount to dehumanisation. The use of a slur may share some features with dehumanisation, but is not itself dehumanising.

One might object that, in the examples above, we have wilfully blurred the line between the justified moral inferiority of an individual and the unjustified inferiority of a whole group.[3] We can rightly rank individuals based on their moral character. But by assumption, we can't rightly rank whole groups based on some prejudice of ours, like a prejudice against gender, ethnicity, or religion. Slurs are designed to express inferiority of the latter kind. Since the two kinds of inferiority are distinct, whether we can align dehumanisation with one of them shouldn't count as evidence for the possibility of aligning it with the other.

I reply that we are not trying to assimilate the ranking associated with, or introduced by, slurs to the moral ranking. The problem is strictly with the explication of weak dehumanisation in (1). The two kinds of ranking are indisputably different. But they are similar in that they both involve thinking of the target as inferior and unworthy of full respect. Of course, the reasons for thinking so are different in the two cases. Conceivably, that's precisely why we may accept one kind of ranking, but not the other.

The objector might now follow two lines of argument. Both of them lead to the question of respect. On the one hand, it is tempting to tweak Jeshion's formula (1a) just a bit:

(2) *x* thinks of *y* in a dehumanising way in the weak sense iff:
i. *x* thinks of *y* as a human, but an inferior one, and unworthy of full respect, and
ii. *x* doesn't have good reasons for thinking so.

Analogously for (1b). To stick to (1a), since there are good reasons to think of *A* the murderer as an inferior human, thinking of him in such a way doesn't amount to dehumanisation. But this move begs the ques-

at least plausible that you attribute it to character and actions to varying degrees. Whether you are a consequentialist or a deontologist, you are bound to think of an avaricious miser, say, as morally inferior, other things equal, compared to a generous benefactor. See, e.g., Kagan (1998: 78ff) for some discussion. It doesn't follow though, at least according to the deontologist, that a person may have no moral status whatsoever. I return to this crucial issue shortly.

[3] Slurs defined as a special kind of group pejoratives: Diaz-Legaspe (2020: 1412–1413).

tion: *Is* it ever justified to think of anyone as unworthy of full respect? For suppose it isn't. Then the condition (2ii) is vacuously true, and we are back to the same problem as earlier. The answer, therefore, must depend on the concept of respect used in Jeshion's explications.

Alternatively, one might object that, though contempt and holding in low regard are not sufficient for Jeshion's "strong dehumanisation", they are sufficient for weak dehumanisation. Weak and strong dehumanisation, says Jeshion, "align in a continuum": there is no "hard demarcation" between the two notions (Jeshion 2018: 80). At the extreme, we find rare, perfectly illegitimate dehumanisation. But you can dehumanise people to a certain degree when your act is not that rare and is perhaps legitimate. This is reflected in the situations where we say that contempt is "fitting", as may be the case with the ruthless killer and the despicable husband. And this is precisely what the concept of weak dehumanisation is intended to capture.

But is there really continuity between strong and weak dehumanisation? There is one conceptual side issue that bears mentioning. Weak dehumanisation is scalar: when you are dehumanised to this or another degree, we say that you are "weakly" dehumanised to that degree, and that you are regarded or treated as only partly human. Strong dehumanisation is Boolean: to be strongly dehumanised is to completely lack the status of a human. So to be strongly dehumanised is to be completely dehumanised. Putting the two together, we have a problem. You may be lightly or gravely injured, but you can't be "completely" injured. You may be a little dangerous or very dangerous, but not completely so. In the other direction, you may be dead, but not a little dead or very dead. We can't, that is, have it both ways: a concept can't be both scalar and Boolean. Thus, we can't speak of people dehumanised to a degree, but also of (other) people to be completely dehumanised.

It would be unwise to pin the critique on this observation alone. Jeshion might drop the talk of the "continuum" and treat weak and strong dehumanisation as two different concepts, one Boolean and the other scalar.

Hence, the main question is the same as before: what is the concept of respect used in the explications (1)? Textual evidence is inconclusive—indeed, Jeshion says nothing about her understanding of respect. One possibility is that she interprets respect in a broadly Kantian way. "Humanity in the moral sense" is meant to refer not to diverse moral qualities possessed by some humans and not by others, but rather to an intrinsic human quality *F* that separates humans from mere things. Kant's own candidate for *F* is the capacity to set goals for oneself (he also insists that this capacity, when exercised autonomously, converges on the capacity for moral action; but we needn't follow him there). Can you be an inferior or superior human in this sense? You are either able to set goals or not. Even if you don't set them very well, or even if your goals aren't very good, you are still an end-setter. The concept of humanity so understood is not scalar, and it's not meant to be.

Further, as we see in (1a), Jeshion couples the idea of inferior humanity with "unworthiness of full respect". But if we think of moral respect along Kantian lines, then the contrast between "full" respect and "limited" or "partial" respect is impossible to draw. Under a standard interpretation, moral respect is intended to be a normative threshold we shouldn't cross in our treatment of a person, good or bad.[4] Suppose you believe that torture is incompatible with moral respect. This means that no human, no end-setter, good or bad, should be tortured. And if you propose to torture a person just a little, or very seldom, you don't show some imperfect moral respect; rather, you fail to show moral respect altogether.

To put it another way: There are intelligible idioms of ranking people *qua* moral agents and of sorting them into "superior" and "inferior". You may have harsh words to say about a person's moral character, his standing as a moral being. You may, in so doing, also express a lot of contempt toward the person's actions and his character. But neither the idioms of ranking, nor the attitude of contempt so expressed, should have anything to do with moral respect. Vice versa, when we ask whether a certain person (or more precisely, a certain object that may in fact be a person) should be treated with moral respect, we don't mean to enquire *how much* respect is due to him, and how superior or inferior this person is.

To return quickly to (2), the rationale behind it is to look at *A,* the murderer, say, as unworthy of full respect, as a human, but an inferior one. But we now see that the attitude of respect does not scale, and that even a murderer deserves full respect. That is, the condition (2ii) is indeed vacuous, and the previous objection holds.

To prevent a misunderstanding: I'm not arguing, absurdly, that because all persons deserve respect, you can't, in practice, *de facto* deny them respect partially or wholly. I am arguing that, because of how the concept of moral respect is understood, it is conceptually impossible to deny people respect *to a certain degree*. It is impossible to say, in this broadly Kantian sense, that *A* treats *B* as worthy of some respect, but not of "full" respect. Because of the original Kantian contrast between respect and honour, respect is a matter of all or nothing.

This last claim isn't rigidly aligned with the standard interpretation of Kantian texts mentioned earlier. The same contrast between scalar and Boolean concepts is available in the accounts intended to deviate from Kant. According to Stephen Darwall, appraisal respect (meant to model "honour") may be given to a smaller or greater degree. Recognition respect (meant to model "moral respect") is "acknowledgement" of the person's dignity, the disposition to give at least some weight to the person's preferences in your practical deliberation. Self-described non-

[4] The "standard" interpretation: e.g., Hill (1991: 11), Wood (1999: 135).

Kantians like Sarah Buss and Joseph Raz defend the notions of respect broadly in line with this formulation.[5]

It might be retorted, and I encountered this idea in discussions, that Jeshion's account is meant to describe bigoted thought, and that if such thought is shown to be incoherent, so be it. Hence, slur-users *do* dehumanise their targets, though they do that incoherently. But this response goes against some very basic demands of charity. It is not as though slur-users freely admit to dehumanising their targets. Nor does the idiom of dehumanisation belong in their vocabulary. *We* impute the concept of dehumanisation to them. By doing so, we are supposed to make certain moral qualities of their linguistic use intelligible. Therefore, if the concept of dehumanisation is shown to be incoherent, this is a failure on our part as interpreters of bigoted speech and thought. It is a sign of our failure to impute, coherently, a dehumanising thought to slur-users.

Another line of defence of Jeshion's account might run as follows. We should read the preceding argument as a *modus tollens*. Since Jeshion's account entails that respect is scalar, the notion of respect she is using is not remotely Kantian. It is some other notion of respect. Unlike Kantian respect, it is not a matter of all or nothing: it *is* scalar. Thus, consider an idea, floated by several authors, that respect is a form of care and considerateness for a particular person, sensitivity to his needs and preferences.[6] Call this "c-respect". Let *l* be an interval within which it is acceptable to treat you with care. Appropriate care falls within *l*. There is, however, a certain threshold I shouldn't cross. If I give you too little care, let alone positively disregard your preferences, that would be unacceptable. Accorded c-respect within *l* is strictly normative, and a failure is a moral violation. By the same token, there are degrees of violation: we can recognise mild c-disrespect and gross c-disrespect. To put this all together, when you use a slur, you weakly dehumanise me, and when you do so, you also show me some measure of c-disrespect. We associate slur-use with weak dehumanisation, and we explicate the latter through c-disrespect.

The notion of c-respect, I agree, is a coherent one capturing some intuitions we have about "respect". But how can the use of slurs be a show of c-disrespect? How can we fail to care for people by using a slur? The only plausible answer, I think, is that slurs are offensive and insulting. Thus, according to Dillon, one way to show respect to people is by "praising and honoring" them (Dillon 1992: 109). Correspondingly, we fail to show respect when we "scorn" or "contemn" them, or when we act "arrogantly" towards them. Furthermore, it is also plausible that insult and offence come in degrees. They allow for differential moral valuation depending on their severity: slurs offend, and different slurs

[5] Semi-Kantian respect: Darwall (2006: 122–123), Darwall (2013: 18–19, 264–270), Buss (1999: 536), Raz (2001: 159).

[6] See, e.g., Dillon (1992), Ullmann-Margalit (2011), Maclagan (1960).

offend differently. This feeds into the proposal that, just like weak dehumanisation, c-respect is scalar.

On this reading, then, Jeshion merely wishes to point out that, since each use of a slur is mildly or grossly offensive, it is for this reason also mildly or grossly dehumanising (in the weak sense of it). There are two obvious problems with this interpretation. First, it conflicts with Jeshion's own claim, noted above, that dehumanisation is in a different league from "derogation" and "debasement" that presumably characterise insult and offence (Jeshion 2018: 80).[7] Secondly, I take it that the rationale of linking dehumanisation to slur-use is in marking out a uniquely problematic moral aspect of slurs. Thus, ordinary offences are not supposed to dehumanise. If they were, there would be nothing special about slurs.

If dehumanisation is spelled out in terms of c-respect, however, then every offence or insult would also dehumanise, at least to some degree. Hence, slurs should lose their (im)moral uniqueness. This result would clash with the whole dialectic of Jeshion's argument. And therefore, it is not plausible to link the explications in (1) to the notion of c-respect.

I conclude that the notion of weak dehumanisation is either outright incoherent (if linked to Kantian respect) or does not serve Jeshion's purposes (if linked to c-respect). There is perhaps some other notion of respect that can be used to avoid both pitfalls. Since I can't think of any, and since Jeshion's discussion provides no clues, I won't pursue this possibility.

Can't we, finally, jettison "weak dehumanisation" altogether and stay only with the "strong dehumanisation" explicated in (1c)–(1d)? To strongly dehumanise a person in thought is to think of him as a subhuman, and to strongly dehumanise him in action is to treat him as a subhuman deserving no moral respect whatsoever. Yet this is just too strong for slurs. A speaker may use a slur and, at the same time, predicate a quality of the victim compatible with the affirmation of moral respect.[8] The generics in (3), prejudiced and misleading as they are, are intended to attribute some bad traits to the targeted groups:

(3) a. Wops are sleazy, but they have a great artistic taste.
b. Chinks have no taste, but they are highly intelligent.

If these slurs dehumanise strongly, then the bigoted speakers in (3) mean to deny the slurred groups a *minimally* respectful treatment accorded to humans. That's highly implausible. It's hard to believe that those speakers are prepared to accord to typical Chinese or Italians the kind of treatment we see in concentration camps, for instance. It is similarly unlikely that the speakers in (3) even think of these groups as separate categories of non-humans or subhumans. Indeed, the mem-

[7] The same idea underlies Jeshion's appeal to the distinction between "appraisal respect" and "recognition respect" (Jeshion 2018: 84).

[8] The argument here echoes the one in Camp (2013) and Anderson and Lepore (2013) against the association of slurs with contempt. It is addressed in Jeshion (2018: 85–86, 96).

bers of these groups are attributed the qualities, such as taste or high intelligence, that are often thought to be uniquely human. And since the speakers profess a dislike of "wops" and "chinks" in some respects and admiration in others, it is not even clear that they mean to deny honour to members of these groups. (3) may be read as little more than expressions of a grudging admiration mixed with hostility or envy.

At this point, one might argue that the speakers in (3) could still be "quintessentialists" in the sense of Leslie (2013, 2017). They might essentialise their targets by attributing to "wops", say, a hidden, not necessarily biological substance ("quintessence") causally responsible for their sleaziness, i.e., "their most important, stable, and enduring" property (Leslie 2013: 110). This tendency to quintessentialise is sustained by their use of nouns and generics (Leslie 2017: 418). And it is also plausible that their quintessentialist tendency is part of the explanation why those speakers continue to entertain prejudices toward Chinese or Italians, and why they are amenable to the introduction of slurs for these groups (we don't have to assume, though, that the extension of "wop" and "Italian" must be the same). By contrast, they don't derogate or discriminate against accountants, to use Leslie's own example. One reason for that is that they don't think there is any accountant-quintessence.

Well, conceivably, the bigots in (3) do quintessentialise Italians and Chinese. But we can hardly infer from that any fact of strong dehumanisation.

For chances are that these bigots, in fact, along with us, quintessentialise many other groups identified through the same mechanism of noun use. Thus, we hear them say, "He is just a lawyer!" So, we expect them to entertain some negative attitudes, likely prejudices, toward lawyers *qua* lawyers. Unlike accountants, lawyers have quintessences, or so our bigots think. Other times, the same bigots declare, "He is a human, not a chatbot!" They mean to say that something uniquely human, the human quintessence, explains human emotional responses, say. Our bigots may well believe that. But of course, it doesn't follow that they are in any way committed to denying lawyers or humans minimal respect. Psychological essentialism does not constitute strong dehumanisation, though it may certainly be a small step towards it, alongside many other factors.[9]

III

In this section, I examine another sophisticated attempt to tie slurs to dehumanisation that, however, is markedly different from Jeshion's.[10] At the end of §II, I appealed to the observation that Jeshion's view fails

[9] I explore these other factors in §IV.

[10] Neufeld (2019, 2020) offers another interesting argument that I won't be able to address here for the lack of space. Neufeld ties the use of slurs to essentialism, essentialism to the denial of free agency, which is in turn linked to dehumanisation. For an instructive early argument in the same direction see Clor (1969).

to do justice to the use of ordinary slurs like "wop" or "chink". That was fair: according to Jeshion, all uses of slurs, certainly uses of racial slurs, are supposed to dehumanise. But what if only a few select slurs dehumanise, while others don't?

A version of this view is defended at great length in Tirrell (2012). Taking the Rwandan genocide as her case study, she argues that certain slurs dehumanise when they are "deeply derogatory" (call such slurs "deep slurs"). Compared to ordinary slurs, deep slurs have five unique characteristics (Tirrell 2012: 190–193). Let me mention two. Deep slurs impute certain essences to the individuals designated by the term (i.e., the targets of the slurs). As observed earlier, it is questionable whether the users of "wop" or "papist" would seriously contemplate an essence common to all Italians or all Catholics. But, we now speculate, to use a slur like "*inyenzi*" ("cockroach") or "*inzoka*" ("snake") is precisely to attribute an essence, presumably a biological one, to the individual. Secondly, a slur like "wop", at least in categorical subject-predicate sentences, may reliably be used to express some negative attitude toward Italians. Deep slurs, on the other hand, are not primarily vehicles for attitude expression. The presumptive deep slur, like "*inzoka*", is "action-engendering" in the way that "wop" is not (Tirrell 2012: 176, 195). Multiple social functions are central to the use of deep slurs: "they are used to dominate, demean, or dehumanize people" (Tirrell 2012: 192). To use a deep slur is to specify which treatment is appropriate for the designated person. Moreover, sometimes the slur is used to "assign" a particular social role to the target (Tirrell 2012: 193).[11]

Let's try to unpack these claims. To stick to our main concern here, how exactly can we use deep slurs to dehumanise others? The answer is unusually complex and demands an extraordinary exegetical effort. Sometimes Tirrell explains the link to dehumanisation by downplaying the distinction between language and behaviour (Tirrell 2012: 206). As Tirrell says repeatedly, to use a slur is to "engender" action. It is a form of "linguistic violence" (Tirrell 2012: 176). All the same, there are two different ways for the slurs to engender action. At an early, "preparatory phase" of the genocide, the use of deep slurs "prepares" the society for physical violence (Tirrell 2012: 190, 200–201, 203, 205). More explicitly:

(4) At an early stage of the genocide, if S utters "x is a cockroach ("*inyenzi*")", then: (i) S is committed to the assertions "x is a pest", "x is dirty", "x is hard to kill", "x should be killed", and other; (ii) S is committed to the permission of the treatment of x

[11] Interestingly, this claim is also endorsed by Jeshion who applies it to slurs across the board, however: "By using the slur ["kike"], and by virtue of its semantics, the anti-semite indicates that being Jewish is a *fundamental negative characteristic-defining feature* of the target's identity *qua* person. As such, the speaker uses the slur to frame and dictate how others ought to conceive of and treat its target. If the use of the slur is public, the speaker invariably performs an act that contributes to making the property part of the target's *social identity*." (Jeshion 2018: 83, her italics) See also footnote 13 below.

as a pest, of the killing of *x*, and of certain other forms of treatment.

At this stage, linguistic use is still "dissociated from social behavior" (Tirrell 2012: 204). Let us say that at this stage, slurs have a "causal role" in dehumanisation.

The situation is different, Tirrell argues, at a later stage, when dehumanisation is accomplished by assigning particular social positions (Searlean status-functions) to the targets.

Specifically, deep slurs like "*inyenzi*" are used to assign definitive status-functions to the targets (Tirrell 2012: 196, 213). Hence, slur-use is then directly associated with harmful treatment like depriving the target of human rights. Let us say that at a later stage, slurs have a "constitutive role" in dehumanisation. Slur-use "constitutes" genocidal acts (Tirrell 2012: 185), rather than merely causing them. We may summarise this idea in the following clause:

(5) At a later stage of the genocide: If *S* utters "*x* is an *inyenzi*", then *S* assigns to *x* a certain social status such that it is permissible to kill every individual having that status (same for "*inzoka*").

We have, therefore, two different lines of argument here. Slurs dehumanise at an earlier stage of genocide by training the future perpetrators to see the victims as subhuman, and to facilitate their extermination. Slurs dehumanise at a later stage of the genocide by classifying their victims in such a way as to make them directly liable to extermination. Yet I will now argue that neither of these two arguments withstands scrutiny.

The constitutive role of slurs in dehumanisation. Clause (5) may be illustrated by the case of Simon Bikindi (214–215). At the height of the Rwandan genocide, a Hutu political agitator is visiting a prison where he identifies several Tutsis *as* Tutsis, who are then instantly killed. The treatment inflicted on the Tutsi victims is arguably dehumanising. I will not pause to ask why we judge it so. I will instead assume that a denial of basic legal and social rights to a person *is* a paradigmatic case of a dehumanising treatment.

Such a change in the legal-political status of a person may be described in a Searlean fashion as an assignment of status-functions. And these status-functions, we may agree, are dehumanising in the sense of denying minimal social roles to the victims. Yet, what does the use of slurs have to do with this? Very little. The treatment of Tutsi prisoners was dehumanising, since they lost any political status, their preferences were completely ignored, their lives counted for nothing. But addressing them with slurs did not *subject* them to that treatment. In fact, Bikindi, according to the report cited by Tirrell, didn't even use slurs. All he had to do was to identify the prisoners as Tutsis in some way, which is indeed what he did. Of course, he could use the slurs "*inzoka*" or "*inyenzi*" instead. But their only role in the ensuing dehumanising treatment—i.e., the brutal killing devoid of any pretence of legality—was identification. The persecutors had to know who the

Tutsis were, and "*inzoka*" and "*inyenzi*" could fulfill that function well. There was nothing special about them in that regard; however, other expressions, gestures, and all sorts of indirect speech acts could do the same.

You might say: this misses the point, for how were the status-functions assigned in the first place? Didn't slurs play a critical role in classifying Tutsis as subhuman and thus enabling their dehumanising treatment? Beyond mere identification, slurs have an action-engendering normative force: they "generate permissions" for non-linguistic behaviour (Tirrell 2012: 176). Therefore, when you say, "*A* is an *inyenzi*", don't you classify *A* as someone who should be killed with impunity?

I reply that the dual-content function of the slurs may be plausible. At any rate, let's grant it: slurs serve to both identify and classify. Still, no dehumanising *treatment* should follow, nor is it even remotely likely to do so. When I say, "This is my house", pointing at Blenheim Palace, I classify it as my house and lay my claim to it (according to Searle). But the palace wouldn't become my house following the declaration. What is required, in addition, is making others accept my classification.[12] A host of social and political conditions have to be met for my bare classification to actually assign the intended status-function. Observe that this is a perfectly commonsensical take on what happened in Rwanda, according to Tirrell's own narrative. It was not until 1994 that a declaration such as "*A* is an *inyenzi*" could assign the dehumanising status-function to *A*. That is because by that time, the Hutu population had *accepted* that Tutsis had to be exterminated.[13]

In short, in 1994, deep slurs could be used to identify the victims for the subsequent dehumanising treatment. This could be done by various other linguistic or non-linguistic means. Addressing *A* with a slur did not subject *A* to a dehumanising treatment, did not constitute it, and did not, *by itself*, ensure any such treatment. At times, indeed, Tirrell seems to appreciate the very special social conditions in Rwanda that could plausibly link slurs and speech in general to dehumanisation and genocide. At times, she seems to downplay them entirely. Consider this passage:

In *most* language games that use racist derogations, the derogation is a *function* attached to a person, whose status changes because of the imposition of the function. Racist derogations are status-functions, for they tell the target, "You count as a so-and-so here." (Tirrell 2012: 213, italics added)

These claims are unconvincing. Derogation alone can't assign any status-functions, any more than I can make a palace my house by call-

[12] Declarations and acceptance (Searle 2010: 85).

[13] A parallel complaint should be lodged against Jeshion's version of this idea (footnote 11). How exactly would the use of "kike" meaningfully "contribute" to the target's "social identity"? What would such a contribution consist in, exactly? And wouldn't this require a class of non-trivial social conditions to obtain, rather than just the bare use of a slur?

ing it "my house". For the function to be assigned, it must gain widespread acceptance and become an institutional fact. In most cases of derogation, there isn't any such acceptance.

I conclude that there is no good argument to show the constitutive role of deep slurs. However, a large question still remains, and it is whether their use played a unique causal role in creating the conditions for a future dehumanising treatment. Through their systematic use, for example, prejudices may gradually become entrenched, thus paving the way for dehumanising treatment. This is the question of their role in dehumanising thought at the early stages of the genocide, and I turn to it next.

The causal role of slurs in dehumanisation. The predication "*A* is a snake" may be understood literally or metaphorically. If understood literally, then *A* is said to be a member of the animal kingdom having a particular appearance, mobility, and other physical characteristics, including the manner in which it can be killed most efficiently. If understood metaphorically, *A* may well be human but possess some snake-like characteristics: attacking quickly, escaping dangerous situations, moving stealthily, and poisoning reputations. Only some of these characteristics may be predicated of animal snakes. Others will be only metaphorically similar ("poisoning reputations") to those of animal snakes. At any rate, using this predicate does not license an inference to the conclusion that the metaphorical snake should be literally killed, let alone killed in some specific snake-like way.

If we accept Tirrell's idea in (4) about the link between assertion and action, then the difference between the literal and metaphorical uses may be stated as:

(6) a. If a speaker *S* asserts on the occasion *O* that *x* is metaphorically *F*, then: (i) *S* is committed on *O* to ascribe to *x* some properties *G1*, *G2*,… that are widely ascribed to *F*s in *S*'s linguistic community and are salient on *O*; (ii) *S* is committed to treat *x* the way *G1*s, *G2*s, …are treated in *S*'s community.
b. If a speaker *S* asserts on the occasion *O* that *x* is literally *F*, then: *S* is committed on *O* to treat *x* the way *F*s are treated in *S*'s community.

Now, in the course of the Rwandan genocide, was the use of "*inyenzi*" and "*inzoka*" metaphorical or literal? On the one hand, Tirrell explicitly says that the use of "*inyenzi*" was in fact metaphorical (Tirrell 2012: 200–201; also compare the idiom of "depiction" on page 203). But then the proposed explanation does not add up. To repeat: When you say, "Harry is a pig", your utterance, insulting as it may be, does not commit you to feeding Harry the leftovers or delivering him to the slaughterhouse. Certainly not: Harry is only metaphorically a pig. When, in the Rwandan context, a speaker says, "*A* is an *inyenzi*", *A* is, again, only metaphorically a pest. Yet, according to (4), the utterance inexplicably commits you to killing *A* like any other (literal) pest. This is *ad hoc*. A metaphorical use of the predicate "*x* is an *inyenzi*" should be like any other metaphor. Then its use is no more dehumanising than the use of "*x* is the sun", "*x* is a rat", "*x* is a lion".

Other revealing details in Tirrell's account suggest, however, that at least for some speakers, the use of deep slurs was *not* metaphorical.[14] Consider this crucial passage about how the inference was actually made:

> The extremist propaganda *described* Tutsi as cockroaches or snakes. For many *uneducated peasants*, if the official authorities state that Tutsi are snakes, *it can't be wrong*. If the local official of the commune orders people to kill snakes, *it makes sense*. When you kill a snake, you smash its head, then you cut it up in different places to make sure it's really dead. These very same forms of torture were inflicted on many Tutsi. (Tirrell 2012: 205, citing a report by André Sibomana, italics added)

A plausible reading of this report is that the Hutu peasants understood the predicate "*x* is an *inzoka*" literally. They took it to have plain descriptive content. Because Tutsis were literal, animal snakes, they had to die a snake-death, rather than a more familiar human death. This literal understanding was not doubted despite its evident absurdity, since it came from the authority.

But then the explanatory challenge should lie elsewhere. If I come to believe that you are a fly, don't ask why I spray you with a repellent: that's what people do to flies (rightly or wrongly). The challenge would rather be to explain how I came to have this preposterous belief about you in the first place. In the Rwandan situation, at least if we take Sibomana's report at face value, the challenge is to explain how the peasants could trust the authority so blindly as to develop such a radical delusion. Potentially, the use of slurs could be a causal factor in the treatment of the Tutsi victims if one could demonstrate that it was uniquely suited to trigger the delusion. Of course, Tirrell's argument invokes no such evidence. On the face of it, the Hutu peasants merely understood "*x* is an *inzoka*" as a descriptive term.

Supposing, however, that slurs used by the authority *could* cause that radical delusion, should we *then* describe the resulting behaviour as "dehumanising"? This kind of scenario provokes conflicting intuitions. Compare insults: you mistake your elderly boss for a friend and approach him with "Hey, dude!" The boss may *feel* insulted. But once the facts are known, not only did you not intend to insult him, but there was arguably no insult to begin with.[15]

As I said, intuitions vary. At a minimum, in the Rwanda situation, we should only insist on the idea that dehumanisation, if there is any, does not explain why the killings happened. Furthermore, there is at least one theory of dehumanisation that delivers an unambiguous result. According to Smith (2020, 2021), a radical delusion of the kind

[14] See Smith (2021: 18ff) for an extended statement. I address Smith's view in §IV.

[15] For a useful discussion on insults see Milić (2018).

apparently experienced by the Hutu peasants is, on its own, irrelevant to the fact of dehumanisation. I return to this issue in §IV.

The upshot is a dilemma. If the use of deep slurs was metaphorical, then no specifically non-human properties were attributed to the target. No properties were attributed, at any rate, that could, *in themselves,* amount to thinking about the target in a dehumanising way. Moreover, if there were such properties, then the use of ordinary metaphors would be equally dehumanising. This is a cost we should be unwilling to pay. If, by contrast, the use of those predicates was literal, then there was a tragic delusion. Hutu peasants killed their victims, since they came to think of them as real, animal vermin and real snakes. Here we can explain the treatment of the victims, but so far as it was a result of a delusion, dehumanisation, if there was any, remains explanatorily idle.

Now, if we embrace the first horn of the dilemma, we must say that the Hutu speakers did not at all think of the targets as subhuman or non-human. If we embrace the second horn of the dilemma, we must say that the speakers did think of them as non-human, but only because of a delusion prompted by their trust in the authority. This explanation, although possible, does not generalise and is plainly irrelevant to our purposes. Ultimately, therefore, Tirrell's account doesn't help to align deep slurs either with dehumanising thought (at an early stage of the genocide) or with dehumanising treatment (at a later stage of the genocide).

IV

So far, my discussion has been negative. I have argued that the views examined in §§II–III fail to offer good reasons for either a constitutive or a causal link between the use of slurs and dehumanisation. It would be fitting to conclude with an argument that shows, in principle, why we can't find such links. One way to proceed is to offer a theory of slurs' meaning and then explain why slurs have little to do with dehumanisation. This I can't do: given the very large literature on both subjects, it would at least require another long paper. For the same reason, I won't be able to offer a theory of dehumanisation. Instead, I will claim that, within the debate on slurs, the notion of dehumanisation was invoked for certain rhetorical purposes. But the notion that could serve those purposes cannot plausibly be aligned with the use of slurs, whatever our position on their semantics and pragmatics.

Why would slur theorists insist that the use of slurs is dehumanising? What was their reason for using the idiom of "dehumanisation" specifically? Call this broad claim the "dehumanisation thesis". Conceivably, their motivation was to charge slur-users with a grave moral violation. This goes pretty much without saying, and explicit statements are hard to find. Still, recall Jeshion's account in §II. Jeshion asks whether terms like "derogate" or "degrade" could adequately describe the linguistic function of slurs. In her view, a slur-user merely regards the target of slurs as an inferior human, rather than downright subhuman (the case of strong dehumanisation). But precisely because derogation and dehumanisation are "continuous" with each other, we must insist on the idiom of "weak dehumanisation" (Jeshion 2018: 80). In general, the writers we have discussed wouldn't have invested a major effort in showing that slurs dehumanise, had they not assumed, and wished their readers to assume, that dehumanisation is a very serious moral wrong. Call this the "moralistic assumption".

Hence, the concept that slur theorists are interested in (call it "m-dehumanisation") must satisfy this condition:

(Gravity) M-dehumanisation is a moral violation of the highest degree.

Its next feature is a corollary of the first one (already noted in §II):

(Rarity) M-dehumanisation is relatively rare.

A grave moral transgression can't happen too often. How often is too often? Plainly, no general formula is possible. Still, a grave transgression can't be routine. For example, some authors believe that many phenomena of modern culture are dehumanising. In Montagu and Matson, dehumanisation is interpreted as "the victory of the mechanical equation over the organic principle" (Montagu and Matson 1983: xxix). As a result, people are "emptied out of essential humanity—in order to be restocked with artificial needs and scientifically conditioned reflexes." Punk rock, Martin Scorsese's movies, professional sports, Marcel Duchamp's art—all of them dehumanise by representing people as, or encouraging them to be, mechanical and animalistic. But even though these authors diagnose certain social and cultural activities as a malaise, they don't intend to stop or ban them. Clearly, they don't condemn artists or athletes as perpetrators of a terrible moral transgression.

Thus, if nearly all areas and cultural processes of modern life are dehumanising, as per Montagu and Matson, then this fact alone should diminish the scale of their putative immorality. Certainly, there will be conditions, as in totalitarian and genocidal regimes, that would enable a routine perpetration of moral crimes. But then, in the larger historical context, these conditions should themselves be relatively rare. If, for example, you are told that all of humanity is stained by an original sin, then the immorality of that bare sin can't be too high.

Similar considerations favour another feature:

(Complexity) M-dehumanisation is a relatively complex task to accomplish.

Again, we can't tell in advance how complex it is. But we may agree that something as simple as moving your finger can't m-dehumanise. Easily performed actions would violate (Rarity).

If such are the constraints on the concept of m-dehumanisation, what are the implications? Well, in the first place, there can't be m-dehumanising thoughts. They can't plausibly meet any of the three constraints. No bare thought, however morally repugnant, should count as a *grave* moral transgression. The same applies to rarity and complexity. Thus, from the outset, only a specific treatment could be considered m-dehumanising. And if the use of slurs does not constitute such a treatment, then it can't properly be described as "m-dehumanising".

This conclusion shouldn't be too controversial. Paradigmatic instances of dehumanisation that satisfy our three criteria would be certain kinds of treatment inflicted on humans, like imprisonment in labour camps, genocidal violence, and possibly some instances of extreme poverty.

Conceivably, that is why dehumanisation is associated with a grave moral wrong to begin with. If the present view is at all controversial, it is only because we now say that these forms of treatment are, in fact, the *only* proper instances of dehumanisation (interpreted as m-dehumanisation).

Perhaps, however, our imputation of the moralistic assumption is uncharitable, and the dehumanisation thesis has some other reasons behind it. Perhaps the authors we have explored here claim that slurs dehumanise only to highlight specific qualities of thought characteristic of slur-users. Further, they are entitled to call those qualities "dehumanising" not because of their immorality, but rather because the same qualities are recognised as dehumanising in other areas of research, in particular in social psychology, where the focus is on "dehumanising perception". As before (§III), the claim may be either about all slurs or just some of them:

(7) a. For every slur s, for every speaker x, if x competently uses s, then x has a dehumanising perception of the target of s.
b. There is a slur s such that, for every speaker x, if x competently uses s, then x has a dehumanising perception of the target of s.

There are now two approaches in the psychological literature. According to one approach, primarily associated with David Livingstone Smith, dehumanising perception is Boolean. It has the following condition:

(8) A has a dehumanising attitude toward B iff A sees B as a subhuman animal. (Smith 2020: 19)

As Smith argues at length, dehumanisation is a unique phenomenon (Smith 2021: 26ff). It is not continuous with other derogatory attitudes. In particular, to think of x as subhuman is not to attribute to x a certain descriptive property, like the lack of rationality or of emotions. It is rather to stipulate that x is not "one of us": having first stipulated that he belongs to a natural kind of "humans", the speaker next stipulates that x does not belong to that kind (Smith 2021: 155). This entails, Smith further argues, that dehumanising thought is politically significant (Smith 2021: 157). It is part of the "ideology" designed to "produce, perpetuate, and augment oppression" (Smith 2021: 158).

One immediate benefit of Smith's view is that it resolves the tension surrounding radical delusion (§III). As a result of their (putative) delusion, the Hutu peasants failed to see their Tutsi victims as human at all. And they thought so precisely because they attributed to them a certain descriptive property of being non-human, on a par with the property that snakes or insects might have. In contrast to the properly dehumanising thought, they didn't form that belief on the grounds of first thinking of Tutsis as "not the same as us". Further, their delusion itself was not part of what Smith calls "ideology"—though it was very likely a consequence of that oppressive ideology promoted by the authorities they trusted.

It seems now that only (7b) is a live option: only some slurs, like animalistic slurs "*inzoka*" or "*Judensau*" (discussed by Smith), may be associated with dehumanisation. Yet even (7b) may be too strong. It is a claim about the semantics of animalistic slurs. On many accounts, it should be possible to use a slur competently without entertaining anything like a dehumanising thought in (8).

In fact, Smith himself is apt to reject (7b). The dehumanising thought associated with the slur *s* is part of ideology. But Smith argues that ideologies can remain latent for a long time. This can mean, for example, that at some point in medieval times, the slur "*Judensau*" could be used as an instrument of oppression (to mark out those who are not part of society), given the political conditions at the time and the willingness of the population to persecute the Jews. Later on, as conditions change, the same slur may be used when it is not, in fact, part of the oppressive ideology.

As the conditions change further, the slur can once again become part of the ideology. Then, only for those people who endorse this ideology and participate in its oppressive practices, the use of that slur should be associated with dehumanisation. To give a not-too-unrealistic example, imagine a German émigré in California in 1939 who disapproves of Nazi racial persecution. The émigré has his prejudices: in fact, he is antisemitic. Yet he thinks that no dislike of the Jews, however plausible, should be turned into a state policy. On the lips of this person, the use of "*Judensau*" will not be dehumanising. At the very same time, on the lips of a Nazi official fully on board with racial persecution, its use *will* be dehumanising.

If this reasoning is sound by Smith's lights, as I think it is, then instead of (7b) we should only accept its watered-down version:

(9) There is a slur *s*, there is an occasion *o*, there is a speaker *x*, such that if *x* competently uses *s* on *o*, then *x* has a dehumanising perception of the target of *s*.

The occasion *o* is special: many political factors must be in place to enable exclusion and persecution of *s*'s target. Since the bare use of a slur does not create an ideology and persecution, it does not warrant an ascription of a dehumanising perception. Hence, Smith's account can't be a good reason for the dehumanisation thesis: it warrants a link between slurs and dehumanising perception only on some rare occasions, and only for some slurs. Thus, in the case examined by Tirrell, we can certainly attribute dehumanising perceptions to Simon Bikindi and his ilk. But using "*inzoka*" and "*inyenzi*" is just one way to express these perceptions, and it is not a necessary condition for these perceptions to arise. Equally, as already argued earlier, dehumanising treatment does not require the use of any dedicated slurring terms.

As far as the theory of slurs is concerned, Smith's contention comes down to this: combined with many other political factors, dehumanising thought interpreted along the lines of (8) enables atrocities and violent persecution. Some of these actions and policies we describe as "dehumanising". Slurs, however, have a marginal role to play there. Only under very specific, historically unique conditions can we attribute dehumanising thought to its users. In contrast to Tirrell's view, the use of slurs can't constitute a dehumanising treatment, either. I conclude that Smith's account doesn't give us a good reason for any interesting link between slurs and dehumanisation.

Smith's view is an outlier. Most researchers in this debate favour a scalar notion of dehumanisation. There are several versions of this idea. Haslam (2006, 2020) suggested a "dual model" of dehumanising attitudes. Humanness should be interpreted along two dimensions. Cognitive abilities and self-control distinguish humans from animals. Sympathy, warmth, and emotions in general distinguish us from machines and robots. Thus, dehumanisation in its extreme forms is "animalistic". It is characterised by a vertical comparison: humans are higher, better creatures than the animal-like subhumans. It is also aligned naturally with disgust and contempt (Haslam 2006: 258). "Mechanistic" dehumanisation, in which certain groups are portrayed as machines, is milder. It is characterised by a horizontal comparison: machine-like individuals are seen as strange, emotionally deficient, but not necessarily inferior. Hence, you might appreciate the ruthless efficiency of a "mechanical" person. The effect typically triggered by this form of dehumanisation is indifference, rather than disgust.

A similar idea is in the "stereotype content model" of cognition in Fiske (2011, 2021).[16] Let's examine it in more detail. In their interactions, we are told, people divide social groups along the dimensions of competence, on the one hand, and trustworthiness and friendliness (or "warmth"), on the other. One's own in-group is usually praised both as competent and trustworthy. Elderly people are pitied as incompetent, but trustworthy. Successful out-group members (e.g., Asians in the US) are envied as competent but untrustworthy. Finally, homeless people, members of the "extreme out-group", are perceived as disgusting and are scorned as both incompetent and untrustworthy.

[16] Another closely related idea is developed in the theories of "infrahumanisation", where A is said to infrahumanise B iff A thinks of B as lacking secondary emotions like remorse, pride, or nostalgia, that, according to A, characterise humans. See Leyens et al. (2007) and Rodríguez-Pérez and Betancor (2023) for a review. Although the terminology of "infrahumanisation" is overtly chosen to contrast it with "dehumanisation", the comments I make below about Fiske's view apply to infrahumanisation theories as well.

This model fits very naturally with the idea that slurs are linked to dehumanisation. We can, for example, paraphrase Jeshion's conditions (1) and say:

(10) x targets y with a slur s iff x thinks of y as belonging to an out-group G such that members of G are lacking either in competence or in trustworthiness, or both.

Each slur will have an identification function: it will assign y to a particular out-group G. The right-hand side of the biconditional is what, on this view, constitutes dehumanisation. And, to repeat, when we say that it does that, we are not spelling out a philosophical analysis of dehumanisation. Rather, we simply borrow idioms from the psychological research, like Fiske's.

The problem with this manoeuvre is that Fiske's model seems motivated by the same moralistic assumption we have criticised earlier. That people divide social groups in accordance with certain (alleged) properties is an intriguing generalisation ostensibly supported by empirical evidence. But, Fiske insists further, the categorisations of out-groups are also "dehumanising". Individuals in the extreme out-group are "extremely" dehumanised, the rest are dehumanised to a lesser extent. This is because to dehumanise is to "refuse to acknowledge the complex human experience of other people" (Fiske 2011: 21). To different degrees, we deny them warmth, familiarity, emotions, articulate language, and complex minds.

Note the implications. If you faithfully and patiently care for your elderly grandfather, whom you categorise as "elderly" and who, you believe, no longer has the best command of his mental faculties, then your attitudes are now said to be "dehumanising", at least to a degree. Is there anything wrong with this description? Or rather: why should we choose exactly the idiom of dehumanisation to describe those kinds of attitudes?

I suspect that the choice, again, has to do with rhetoric. We begin with instances of dehumanising extreme out-groups that, on the face of it, are plausibly associated with a moral violation. Then we insist on the similarities between the attitudes toward these groups and the much broader attitudes toward many other groups. Many more attitudes, and many more people, are now implicated in essentially the same kind of violation. The risk here is not only the questionable expansion of moral violations. It is also that the paradigmatic, really serious violation is no longer grave or unique. For nearly everybody is doing the *same thing*, we say, just to another degree.[17]

What of the scorn and disgust toward extreme out-groups, the supposed primary locus of dehumanisation? We are told that:

> Dehumanized perception is a cognitive bias characterized by *spontaneous* failure to think about mental contents—thoughts and feelings in a social target's mind. […] People *spontaneously* fail to consider another person's

[17] For a similar complaint see Bloom (2022).

> mind if that person elicits predominantly disgust, a phenomenon we characterize as dehumanized perception. [P]articipants *spontaneously* think about the minds of dehumanized targets less often than the minds of other social targets. (Harris and Fiske 2011: 175, 180, italics added)

Suppose that this cognitive bias is real. Should we call it "dehumanising"? If this is just another label, it's as good as any. Yet it can't be just another label: its use is intended to strike a chord. We are expected to agree that there is something deeply repugnant about that cognitive failure. This is far from clear. The problem is not only that we should hesitate to judge a "spontaneous" failure, over which a person has no immediate control, as immoral. It is also that merely having certain experiences, let alone spontaneous ones, is not a moral failure. Or if it is, we can't grant it without further argument.[18] It would be different if, in addition, we were shown how these experiences lead to an "m-dehumanising" treatment—that is, to a treatment that can straightforwardly be interpreted as morally repugnant.

I think no evidence for this causal link is forthcoming. For we are not looking for *some* causal link between spontaneous biases of the kind described by Fiske and m-dehumanising treatments like various kinds of persecution. The truth is that many causal factors enable these treatments. Consider, for example, the Holocaust. In the first place, there were transparent political and economic factors that provided ample motivation for murderous actions.[19] Secondly, as many historians have argued, the Holocaust was enabled in large part by the efficient German bureaucracy, as well as the German military and its highly distinctive ethos.[20] These social actors played an incomparably more central role in the Holocaust than Alfred Rosenberg's racist ramblings. It would, however, be strange to call those institutions and their practices "m-dehumanising" merely because of their causal role (though they may perhaps be mechanistically dehumanising in Haslam's sense of the term). So the evidence we should look for must show how biased thinking leads directly and inescapably, pretty much on its own, to m-dehumanising treatment. But since this treatment results from a combination of complex social factors, we are unlikely to find such evidence.

Summing up, there is no good reason to characterise the use of slurs as a form of dehumanisation. A likely explanation for why different writers continue to insist on linking slurs to dehumanisation is their attempt to charge the users of slurs with a grave moral violation (the

[18] For an extended argument why thinking can't be morally wrong, see Sher (2021).

[19] Typical economic factors: Evans (2004: 22ff). Conspiracy theories: Evans (2009: 242ff).

[20] German bureaucracy and the Holocaust: Hilberg (2003: 1059ff). German military and the Holocaust: Elias (1996). The complexity of the more remote causes of the Holocaust, like the tendency to blind obedience, is addressed in Evans (1987: 156ff).

"moralistic assumption"). But whatever moral violation can plausibly be aligned with the use of slurs, it is simply not grave enough to count as dehumanisation. The less likely explanation is that those writers merely aim to join ranks with social psychologists who have developed theories of "dehumanising perception". But in this case, either those theories, like Smith's, are not altogether relevant to the use of slurs, or else they rest on the same problematic moralistic assumption.[21]

References

Anderson, L. and E. Lepore. 2013. "Slurring words." *Noûs* 47 (1): 25–48.

Bloom, P. 2022. "If everything is dehumanization, then nothing is." *Trends in Cognitive Sciences* 26 (7): 539.

Buss, S. 1999. "Respect for persons." *Canadian Journal of Philosophy* 29 (4): 517–550.

Camp, E. 2013. "Slurring perspectives." *Analytic Philosophy* 54 (3): 330–349.

Clor, H. M. 1969. *Obscenity and Public Morality*. Chicago: University of Chicago Press.

Darwall, S. L. 2006. *The Second-Person Standpoint*. Cambridge: Harvard University Press.

Darwall, S. L. 2013. *Honor, History, and Relationship*. New York: Oxford University Press.

Diaz-Legaspe, J. 2020. "What is a slur?" *Philosophical Studies* 177 (5): 1399–1422.

Dillon, R. 1992. "Respect and care." *Canadian Journal of Philosophy* 22 (1): 105–131.

Elias, N. 1996. *The Germans*. New York: Columbia University Press. Edited by M. Schröter. Translated by E. Dunning and S. Mennell.

Evans, R. J. 1987. *Rethinking German History*. London: Allen and Unwin.

Evans, R. J. 2004. *The Coming of the Third Reich*. New York: Penguin Press.

Evans, R. J. 2009. *The Third Reich at War*. London: Penguin Books.

Fiske, S. T. 2011. *Envy Up, Scorn Down*. New York: Russell Sage Foundation.

Fiske, S. T. 2021. "How status and interdependence explain different forms of dehumanization." In M. E. Kronfeldner (ed.). *The Routledge handbook of dehumanization*. London: Routledge.

Harris, L. T. and S. T. Fiske. 2011. "Dehumanized perception." *Journal of Psychology* 219 (3): 175–181.

Haslam, N. 2006. "Dehumanization: An integrative review." *Personality and Social Psychology Review* 10 (3): 252–264.

Haslam, N. 2020. "The many roles of dehumanization in genocide." In L. S. Newman (ed.). *Confronting Humanity at Its Worst*. New York: Oxford University Press, 119–138.

Hilberg, R. 2003. *The Destruction of the European Jews*. New Haven: Yale University Press.

[21] I am grateful to two anonymous referees for helpful comments on an earlier draft of this paper.

Hill, Jr., T. E. 1991. “Servility and self-respect.” In *Autonomy and Self-respect*. Cambridge: Cambridge University Press, 4–18.

Jeshion, R. B. 2018. “Slurs, dehumanization, and the expression of contempt.” In D. Sosa (ed.). *Bad Words*. New York: Oxford University Press, 77–107.

Kagan, S. 1998. *Normative Ethics*. Boulder: Westview Press.

Leslie, S.-J. 2013. “Essence and natural kinds.” In T. Szabó Gendler and J. Hawthorne (eds.). *Oxford Studies in Epistemology*. Oxford: Oxford University Press, 108–165.

Leslie, S.-J. 2017. “The original sin of cognition.” *Journal of Philosophy* 114 (8): 393–421.

Leyens, J.-P., S. Demoulin, J. Vaes, R. Gaunt, and M. P. Paladino. 2007. “Infra-humanization: The wall of group differences.” *Social Issues and Policy Review* 1 (1): 139–172.

Maclagan, W. G. 1960. “Respect for persons as a moral principle: I.” *Philosophy* 35 (134): 193–217.

Mikkola, M. 2016. *The Wrong of Injustice*. New York: Oxford University Press.

Milić, I. 2018. “What counts as an insult?” *Acta Analytica* 33 (4): 539–552.

Montagu, A. and F. Matson. 1983. *The Dehumanization of Man*. New York: McGraw-Hill.

Neufeld, E. 2019. “An essentialist theory of the meaning of slurs.” *Philosophers' Imprint* 19 (35).

Neufeld, E. 2020. “Pornography and dehumanization.” *Australasian Journal of Philosophy* 98 (4): 703–717.

Raz, J. 2001. *Value, Respect, and Attachment*. Cambridge: Cambridge University Press.

Rodríguez-Pérez, A. and V. Betancor. 2023. “Infrahumanization.” *Current Opinion in Behavioral Sciences* 50: 101258.

Searle, J. R. 2010. *Making the Social World*. New York: Oxford University Press.

Sher, G. 2021. *A Wild West of the Mind*. New York: Oxford University Press.

Smith, D. L. 2011. *Less than Human*. New York: St. Martin's Press.

Smith, D. L. 2020. *On Inhumanity*. New York: Oxford University Press.

Smith, D. L. 2021. *Making Monsters*. Cambridge: Harvard University Press.

Tirrell, L. 2012. “Genocidal language games.” In I. Maitra and M. K. McGowan (eds.). *Speech and Harm*. Oxford: Oxford University Press, 174–221.

Ullmann-Margalit, E. 2011. “Considerateness.” *Iyyun* 60: 205–244.

Wood, A. W. 1999. *Kant's Ethical Thought*. Cambridge: Cambridge University Press.

Croatian Journal of Philosophy
Vol. XXVI, No. 77, 2026
https://doi.org/10.52685/cjp.26.77.2
Received: September 29, 2025
Accepted: April 16, 2026

Repeatable Experiences of Music and Texts[1]

CAROLA BARBERO and FABRIZIO CALZAVARINI
University of Turin, Turin, Italy

In The Performance of Reading *(2006), Peter Kivy argues that reading literary texts is performative in nature, like musical performance, where both require active interpretive engagement. This analogy highlights that both activities involve interpretive decisions and can be silently performed, involving an auditory imagery experience of the inner voice. Nevertheless, Kivy argues that, unlike music, reading is not an intrinsic repeatable activity, mainly because the primary motivation for reading novels is to uncover the plot, and once known, the aesthetic pleasure decreases. In the present paper, we challenge Kivy's view with both theoretical and empirical evidence. First, literature, like music, involves cultural traditions of repetition and contains multiple layers of interpretation, similar to how repeated musical performances reveal new insights. Second, existing empirical data show that rereading literary texts, including novels, improves appreciation and comprehension, much like repeated listening to music enhances liking and understanding. Overall, we believe that these considerations suggest that both reading and music allow for repeatable experiences that deepen with each engagement.*

Keywords: Peter Kivy; music; literary texts; repetition.

[1] Although the paper has been conceived and discussed together, Carola Barbero has written section 2 and Fabrizio Calzavarini has written section 3. Section 1 has been jointly written by the two authors. An earlier version of this paper was presented at the conference *Beauty and Change 2025*. We thank the audience of this conference, and in particular Jacopo Frascaroli, Pietro Kobau, Elizabeth Schellekens, and Alberto Voltolini, for their useful comments.

1. Peter Kivy on music and literary texts

1.1 Analogies

The philosophical literature offers several extensive explorations of the analogy between music and language. For instance, Kivy defends the view that music is linguistic because it possesses a syntactic component similar to language, although in an analogical sense: "musical 'syntax' may not be syntax in the literal sense, but syntax-like, as music is language-like, but not language" (Kivy 2007: 216). Music lacks a semantic component – that is, meaning in the literal sense – but can evoke emotive properties that listeners can interpret based on their familiarity with musical structures, thereby forming an emotive "vocabulary" (for discussion, see Di Bona 2019). In the *Performance of Reading* (2006), Kivy further explores the analogy between music and language, specifically focusing on literature. He introduces the claim that the act of reading is performative in nature and that performance is a crucial aspect of understanding the aesthetics of literature. He emphasizes that, just as a musical performance involves an interpreter (the musician), so too does the reading of literature involve the reader's interpretive engagement with the text (for discussion, see Feagin 2008; Ribeiro 2009; Morais 2010, 2017).

Kivy's analogy between reading texts and musical scores highlights several key points. Both activities require individuals to make interpretative decisions, necessitating active engagement to bring the written material to life (Kivy 2006: 67-74). Additionally, both reading and music unfold over time and can be performed silently, critically involving auditory mental imagery (the "inner voice"). Indeed, when reading music silently, skilled musicians "hear" the performance in their heads. Similarly, when reading literature, both expert and ordinary readers can imagine the narrative being spoken, hearing voices, whether the narrator's or the characters' (Kivy 2006: 114–126). In this sense, according to Kivy, "we hear stories in the head, the way Beethoven, when he read the scores of Handel, heard musical performances in the head" (Kivy 2006: 63). An immediate implication of Kivy's view is that congenitally deaf individuals may lack a full aesthetic experience. Since Kivy links the appreciation of literature to auditory mental imagery – such as "hearing" dialogue internally – he concludes that deaf individuals may have a "defective" aesthetic engagement with literary works (Kivy 2006: 68–69).

Interestingly, it has been recently shown that Kivy's phenomenological observations on the analogies between music and texts are at least partially confirmed by empirical data from cognitive neuroscience (for discussion, see Barbero and Calzavarini 2024). Despite being functionally and anatomically dissociated at the cognitive level, as suggested by patient (e.g., Cappelletti et al. 2001; Mendez 2001), eye-tracking (e.g., Cara and Gomez-Vera 2016), and neuroimaging (Mengelli et al.

2017) studies, the two reading experiences involve an auditory simulation of the content, similar to having a performance in the mind's ear. A series of fMRI studies, for example, have demonstrated that regions in the primary and secondary auditory cortices involved in music and voice processing are also activated when both novices and experts are silently reading musical notation and texts (e.g., Yao et al. 2011; Mongelli et al. 2017). This auditory imagery experience seems to be critical for a deep understanding and appreciation of musical scores and literary texts, as suggested by studies on reading abilities in deaf subjects. This reinforces Kivy's analogy between the experiences of music and literary texts.

1.2 Structural disanalogies

Kivy's analogy between music and texts reaches a point of divergence when it comes to the role of repetition. Already in the *Introduction to a Philosophy of Music*, Kivy argued that absolute music – i.e., instrumental music without propositional content – is structurally distinct from literary texts (Kivy 2002: 153–154). According to Kivy, "[…] music, to put it bluntly, is very repetitious. Repetition is an integral part of musical form, both internally and externally" (Kivy 2002: 153). Internally, within the movements of a symphony, musical entities are structured around recurring patterns: "Just as your wallpaper or your rug repeats small designs to make the larger ones, so too, in the movement of a symphony, the internal structure is a structure or repeated […] melodies, melody fragments, or smaller musical motives" (Kivy 2002: 153). Externally, classical musical pieces involve large-scale repetitions, often signalled by musical notations indicating to the performer to play the earlier section again. Kivy observed that these structural repetitions are not incidental but "essential building blocks of most musical forms, during the past 300 years. Musical form, as we now know it, would be impossible without them" (Kivy 2002: 153–154).

In contrast, Kivy argues that literary texts such as stage plays or novels are structurally linear or progressive, avoiding repetition to maintain coherence and forward movement. Kivy highlights the absurdity of imposing repetition into the structure of narrative forms, both internally and externally: "Suppose Hamlet were constructed that way. Then, instead of saying 'To be, or not to be...' once, and then getting on with his life, Hamlet would repeat, every few minutes, 'To be, or not to be...'. Not only that, but each act of Hamlet – it has five! – would be performed twice, the first act being repeated before the second act could be presented, and so on. The absurdity of this procedure hardly needs further comment" (Kivy 2002: 154). Based on these observations, Kivy concludes that music, as "the fine art of repetition", fundamentally diverges from the non-repetitive nature of literary or dramatic texts: "Were narrative fiction as repetitive as absolute music, it would not be the art of narrative fiction as we now have and understand it. And were absolute music as unrepetitive as narrative fiction is, it would not be the art of absolute music as we now have and understand it" (Kivy 2002: 154).

1.3 Experiential disanalogies

The strongest disanalogy between music and texts regarding repetition, according to Kivy, is not structural but experiential: whilst musical pieces can be repeatedly enjoyed and experienced in fresh ways, the same cannot be said for literary texts, especially novels. In *The Performance of Reading,* Kivy argues that the phenomenon of re-experiencing music in the everyday life cannot be entirely explained by the novelty of performance: "For many people have only one recording of each work of music in their record collections, yet they listen to their favourite works over and over again, even when it is the same performance each time" (Kivy 2006: 88). Instead, the phenomenon has deeper roots in the nature of musical experience itself, where aesthetic pleasure does not diminish but instead increases with repeated listening and familiarity. In music, a listener can always derive new interpretations or emotional responses from repeated exposure, even when she is entirely familiar with the piece. Each performance or listening session allows for new insights and pleasures by revealing subtleties in phrasing, openings, dynamics, or structure that were not apparent on first hearing.

On the contrary, Kivy posits that reading, particularly the reading of novels, lacks this same potential for repeated enjoyment as musical pieces once the plot is uncovered. Kivy contends that the primary motivation for reading a novel is to discover its plot, and once that discovery is made, the motivation for rereading diminishes significantly: "To put it crassly, once you know the story – once you know how things come out – the major source of artistic satisfaction has been exhausted. So if, indeed, the normal reader reads a novel a second time, it will be after a period of time long enough for the general outline as well as the details of the plot to have faded from memory, in effect making it as if he or she were experiencing the work for the first time" (Kivy 2006: 84). This is also reflected in our common reading practices, where we are not inclined to revisit novels with the same frequency or enthusiasm as musical works: "There is nothing at all odd in declining to read a novel one has already read, even if it is a masterpiece like *Pride and Prejudice*. Normally, one tends to read a novel but once" (Kivy 2006: 87).

Granted, as observed by Kivy in *Paraphrasing Poetry (for Leasure and Profit),* in case of difficult literary works such as Milton's *Paradise Lost*, or *The Divine Comedy*, or the *Odyssey,* rereading is perfectly possible and legitimate: "A serious, complex poem like *Paradise Lost* has so much going on in it that a reader, even a highly sophisticated one like C. S. Lewis, cannot possibly 'get it all' with one, shall I say, 'readerly' attitude" (Kivy 2011: 373). This is not merely due to the length of such poems, because the same can be said for "many of the sonnets of

Shakespeare, or Donne's notoriously dense and difficult 'metaphysical' and religious poems, brief though they may be in comparison" (Kivy 2011: 373). The critical point is that, according to Kivy, rereading in these cases has little to do with aesthetic pleasure. Rather, the act of rereading is driven by the need to focus on specific aspects for study or intellectual engagement: "In reading such poetry, it is not only perfectly proper but absolutely necessary to read it now for the philosophical, religious, or moral content, or whatever, another time concentrating on this or that aspect of the linguistic or formal structure" (Kivy 2011: 373).

1.4 Kivy vs Nabokov

Kivy's argument stands in contrast to scholars like Vladimir Nabokov, who advocated the art of rereading as the true mark of engagement with literature in its highest sense (the one capable of causing the famous "shiver on the back"). Nabokov famously argued that "one cannot read a book: one can only reread it. A good reader, a major reader, an active and creative reader is a rereader" (Nabokov 1980: 3). In this view, the aesthetic pleasure of reading does not lie merely in discovering the story but in revisiting the text to disclose deeper layers of interpretation. In Nabokov's view, rereading removes the cognitive barriers due to the physical and psychological effort to understand the text, allowing the reader to engage more deeply with the artistic features of the work and to reach a fuller aesthetic appreciation:

> When we read a book for the first time the very process of laboriously moving our eye from left to right, line after line, page after page, this complicated physical work upon the book, the very process of learning in terms of space and time what the book is about, this stands between us and artistic appreciation […]. The element of time does not really enter in a first contact with a painting. In reading a book we must have time to acquaint ourselves with it […]. But at a second, or third, or fourth reading we do, in a sense, behave towards a book as we do towards a painting. (Nabokov 1980: 4)

Yet, Kivy remains sceptical of this argument, asserting that while some readers may indeed find pleasure in rereading, this is not the norm for most literary engagement: "Now it is surely no intent of mine to deny that such novel-reading as Nabokov describes is appropriate and rewarding. But it is not the way most readers enjoy novels, who do enjoy them, including the serious ones, the great ones, the works of genius. And so the above argument is of no avail to me" (Kivy 2006: 89). Accepting Nabokov's argument, according to Kivy, risks overintellectualizing the novel-reading experience:

> The kind of reader [Nabokov] describes is certainly not the normal reader, nor is it the reader or kind of reading for which many if not most of the great as well as the not-so-great novels were written. They were written for a thinking reader, yes. However, they were written to be read and enjoyed and thought about by folks who read for pleasure, read a novel generally only once, and then move on to something else. […] To give the palm only to

> the Nabokov-style reader is, it appears to me, to succumb to a very unpleasant form of intellectual snobbery. The non-repeatability of the novel, then, is, Nabokov's reader to the contrary notwithstanding, the general rule, but, as we have seen, is no argument against silent novel- reading being a kind of performance. (Kivy 2006: 89)

2. On repetition and literary engagement

2.1 Structural analogies

We have seen that Kivy defends a fundamental analogy between music and literary texts, which has also been partially vindicated at the empirical level (for discussion, see Barbero and Calzavarini 2024). Nevertheless, he marked a significant disanalogy when it comes to the role of repetition: unlike music, literature is not inherently a repeatable art form. At the structural level, according to Kivy, literature is not characterized by repetition as an essential feature, whilst repetition is fundamental to the form of music, shaping its patterns and themes. At the experiential level, musical experiences are intrinsically repeatable, for they allow for fresh interpretative engagement each time, whereas literary texts, particularly novels, lose much of their novelty and appeal upon rereading, as the primary motivation of uncovering the plot has already been fulfilled.

In our opinion, however, there are both theoretical and empirical reasons arguing against Kivy's point on the repeatability of music *vs* texts. Let us start with the structural level. Kivy is certainly right in observing that music tends to be a highly repetitive stimulus (for discussion, see Margulis 2013). Interestingly, this is true not only for classical musical pieces, such as symphonies or sonatas, but also for contemporary musical compositions. For instance, a recent study by Yu and Ying (2014) analyses the role of repetitive motifs in Western popular songs from 2000 to 2013, focusing on their frequency, duration, and impact on listening time. The findings reveal that 71.43% of the analysed songs contained at least one repetitive motif, with an average duration of 368.8 centiseconds and an average occurrence of 43.5 times. That said, Kivy fails to observe that repetition is also a critical structural device when it comes to literary texts. As is known, repetition in poetry and in some examples of prose has to do with repeating words, sentences, lines, and stanzas, and can serve to help establish rhyme schemes, meter, and rhythm. These all contribute to the musical quality of poems when they are read aloud.

Clear examples of repetition in poetry include anaphora (repetition of the same word or phrase at the beginning of lines – e.g., Martin Luther King Jr.'s "I Have a Dream" speech), epistrophe (repetition at the end of lines for emphasis), and refrain (recurring lines or phrases, as in ballads or songs). Repetition in poetry serves a variety of functions, contributing to the semantic, aesthetic, emotional, and structural dimensions of the work (for discussion, see, for instance, Ribeiro 2007, 2018; Clarvoe 2009). It is a fundamental poetic device that helps creating rhythm and musicality within the text, reinforces themes and ideas (e.g. in T.S. Eliot's "*The Waste Land*", repetition amplifies themes of desolation and fragmentation), builds emotional intensity (think about Edgar Allan Poe's "*The Raven*", where the repeated word "Nevermore" deepens the speaker's sense of hopelessness), structures the poem, evokes a sense of tradition or ritual, invites participation, explores variation and context, and mirrors natural patterns.

Critically, examples of repetition can also be found in prose, contrary to what Kivy assumes. As observed by Miller, "a long work like a novel is interpreted, by whatever sort of reader, in part through the identification of recurrences and of meanings generated through recurrences" (Miller 1982: 1). Repetition in novels occurs at various levels. On a small scale, verbal elements can be repeated, such as words, figures of speech, or metaphors. On a larger scale, events or scenes might be duplicated within the text. As Miller puts it, "Any novel is a complex tissue of repetitions and of repetitions within repetitions, or of repetitions linked in chain fashion to other repetitions" (Miller 1982: 2).

To make some examples from famous novels, in *The Sound and the Fury* by William Faulkner, repetition reflects Benjy's fragmented mental state, with recurring thoughts like "Caddy smelled like trees" evoking his deep attachment and obsession. Similarly, in Charles Dickens' *A Tale of Two Cities*, the iconic opening line, "It was the best of times, it was the worst of times," underscores the paradoxes of the era while creating a rhythmic cadence that draws readers into the narrative. Toni Morrison's *Beloved* uses repetition to reinforce themes of memory and trauma, as the repeated reference to "124" anchors the story in its exploration of grief and history. In *The Old Man and the Sea*, Ernest Hemingway mirrors Santiago's persistence through repeated descriptions of his struggles, adding a meditative quality to the narrative. In *The Wings of the Dove*, Henry James employs repetition for emphasis and rhythm, such as in Milly's reflection: "And she was dead, dead, dead" (more generally, James in his novels masterfully employs repetition across various levels; for discussion, see Radetič 2018).

These examples demonstrate that repetition in prose, similar to the repetition of refrains, patterns, and motifs in music, is a common feature of literary texts, and interpreting a novel is done in part through noticing such recurrences (Miller 1982).

Indeed, a reader sympathetic with Kivy might object that repetition is an obligatory and architectural feature of musical pieces, whereas in literature, although common, repetition is often local and optional, reflecting a stylistic or rhetorical choice of the author.[2] We believe, however, that this contrast is overstated. First, it should be noted that some musical traditions, such as free improvisation or certain atonal

[2] We thank an anonymous reviewer for this observation.

practices, do not rely on, or at least minimize the use of, architectural repetition. As observed by Gartman, for instance, "[t]he ban on repetition is one of the most important dogmas of free improvisation. This taboo can be explained with the paradox that improvisation means inventing something in the moment, and you can invent something only once" (Gartman 2021: 392). Secondly, even when repetition is structurally critical in music, the apparent difference with literature might be more a question of perceptual immediacy and temporal granularity rather than genuine structural presence. Typically, repetitions in music are easy to individuate for they are compressed in time, differently to literary pieces, where they tend to be more disperse. The patterns of repetition in novels, for instance, such as motifs or thematic and narrative parallels, are often more distributed in time, operating across extended sketches of text. In fact, when repetition in literary pieces is more concentrated, such as in poetry, it is immediately perceived as structural, as we have seen above.

2.2 Experiential analogies

What about the experience of music and texts, which is the main focus of the present paper? Granted, from an intuitive point of view, reading literature is cognitively more demanding than listening to music, given the complexity of language processing, memory load, and interpretative effort. This may help explain why people, in general, tend to relisten to music more easily than re-read novels (perhaps because life is too short to squander such labours). Nevertheless, in our opinion, Kivy's disanalogy, which suggests that repeated encounters with literature are somehow less compelling or fruitful than repeated encounters with music, is challenged by several pieces of philosophical evidence.

First, Kivy's disanalogy overlooks cultural elements of repetition. As observed by Kendall Walton in *Mimesis as Make Believe,* clear examples of cultural text repetition include children's rereading of familiar stories: "it is a notable commonplace that children, far from being bored by familiar stories, frequently beg to hear the same ones over and over again" (Walton 1990: 260). This establishes a parallel with music, where it has been demonstrated that relistening familiar songs has a calming effect on both normal and neurodivergent children (Kim et al. 2024). As again observed by Walton, adult literary culture also supports the value of repetition by explicitly encouraging rereading: "Some adult traditions – ancient Greek theater, Javanese Wayang Kulit – have a relatively fixed repertoire of standard, well-known plots, which nevertheless remain alive and exciting for the audiences" (Walton 1990: 260). This parallels the phenomenon of cultural repetition in musical tradition, where the same compositions were performed and reinterpreted repeatedly, not only in Western classical music (for discussion, see Margulis 2013), but also in Eastern traditions such as Indian music (for discussion, see Jairazbhoy 1971).

Second, Kivy's conviction that our primary motivation for reading a novel is to discover its plot neglects the point that, as Peter Lamarque argues in *The Opacity of Narrative*, literary works are not transparent windows onto their content but are characterized by *opacity*: "Rather than supposing that narrative descriptions are a window through which an independently existing (fictional) world is observed […], we must accept that there is no such transparent glass – only an opaque glass, painted, as it were, with figures seen not through it but in it" (Lamarque 2014: 3). This means that literature can be appreciated for its intrinsic qualities such as its aesthetic, thematic, and imaginative aspects. And this is, of course, what drives our interest when we engage in rereading texts: we do not repeat the experience of reading for uncovering again a story that we already know, but to appreciate the intrinsic artistic features of the text, such as its narrative techniques, style, and the particular way it creates meaning within its literary world. Furthermore, far from being a way of overintellectualizing the experience of reading literature, this is what usually happens in everyday life (Lamarque 2014)[3].

Third, Kivy's disanalogy between literature and music also misses the dynamic, layered nature of literary engagement. Contrary to what Kivy argues, the "meaning" of a literary text, such as a poem, is not "the sum total of the various propositions or 'meanings' the poet intended to convey with his poem" (Kivy 2011: 368). As famously observed by Iser, literature often allows for multiple layers of semantic interpretation, which can be uncovered and appreciated more fully only through repeated readings:

> A second reading of a piece of literature often produces a different impression from the first. The reasons for this may lie in the reader's own change of circumstances, but all the same, the text must be such as to permit this variation […]. The increased information that now overshadows the text provides possibilities of combination that were obscured in the first reading. Familiar occurrences now tend to appear in a new light and seem to be at times corrected, at times enriched. But for all that, nothing is formulated in the text itself; rather, the reader himself produces these innovative readings. (Iser 1984: 10)

This is particularly true when it comes to complex literary texts that include several artistic elements that are not immediately apparent. On a first reading, a reader might grasp the story or basic emotional arc, but on subsequent readings, it is possible to interact with the narrative structure of a text in diverse ways, for instance, by noticing a subtext, a particular character's trait, or a narrative device

[3] Think about poetry: we reread our favourite poems over and over again without feeling bored or tired even if we already know what they are about, and we are eager to repeat that experience. Curiously enough, Kivy's disanalogy is specifically focused on novels and does not take into consideration poetry, which sounds odd because both poems and novels are included in the more general category of literary works as Kivy's admits (Kivy 2006: 101).

(like unreliable narration) that went unnoticed before. For example, modernist literature like the works of James Joyce or Virginia Woolf invites rereading precisely because its narrative style – stream of consciousness and non-linear storytelling – demands active participation and multiple attentive readings. As Joseph Frank convincingly states, "Joyce cannot be read – he can only be reread" (Frank 1968: 19).

Fourth, the active process of uncovering meaning is ongoing and evolves with each rereading. Readers bring their own experiences, knowledge, and evolving interpretations to each new encounter with the text, which allows them to appreciate different aspects over time. As Bence Nanay explains, "We have a different experience [when rereading] because reading heavily relies on the use of mental imagery, and the mental imagery reading conjures up can be very different on two different occasions" (Nanay 2022). When we read, we generate mental images that, by producing internal, perceptual-like experiences involving not only visual images but also sounds, textures, and other sensory details through a multisensory process, bring the text to life, allowing us to "see" and "feel" the characters, settings, and events as if we were witnessing them directly (Nanay 2009, 2016). This imaginative engagement is crucial for shaping our understanding and interpretation of the story while connecting with its emotional content. But mental imagery, as Nanay stresses, is highly variable. It can change based on our mood, prior experiences, or even the specific context in which we are reading. That is why when we reread a novel, the mental images we conjure can differ, leading to a unique reading experience each time. Thus, mental imagery is not a static reproduction of a text but an active, dynamic process, making each reading encounter distinct.

Therefore, the activity of reading, as again observed by Iser, is a "kaleidoscope of perspectives, preintentions, recollections" (Iser 1984: 84), where each sentence creates expectations for the next while also reshaping what has been read:

> Every sentence contains a preview of the next and forms a kind of viewfinder for what is to come: and this in turn changes the "preview" and so becomes a "viewfinder" for what has been read. This whole process represents the fulfillment of the potential, unexpressed reality of the text, but it is to be seen only as a framework for a great variety of means by which the virtual dimension may be brought into being. The process of anticipation and retrospection itself does not by any means develop in a smooth flow. (Iser 1984: 54)

Critically, this explains why, contrary to what Kivy assumes, readers can experience a narrative text differently upon each reading, even if they are already aware of the story, much like how musicians interpret a score differently with each performance.

Indeed, a reader sympathetic with Kivy might still object that there exists a clear asymmetry between music and literature as regards repeated experiences: whilst in music re-exposure typically involves relistening to the entire piece, literary rediscovery is often selective and assisted by memory, since readers usually return to specific key fragments (e.g., openings, lyric highlights), rather than fully rereading. We believe, however, that this asymmetry should not be overstated. First, even in music, repetition does not always amount to full re-exposure: particularly with long and articulated pieces, listeners frequently rely on partial reply, such as relistening the overture of a classical piece (Margulis 2013). Second, the asymmetry might be explained in part by the affordances of the medium: musical experience typically unfolds as a continuous temporal stream, making precise navigation more demanding from a practical point of view and discouraging selective repetition, whilst written texts can be easily navigated non-sequentially, making selective rereading more natural. Critically, the asymmetry does not seem to be stable across reading modalities: audiobooks, in fact, render literary experiences more similar to music in terms of temporal continuity and difficulty of navigation (Rubery 2014). Therefore, the contrast in question seems to reflect more a difference in access conditions rather than a structural distinction in repeatability.

3. The empirical data

3.1 Repeated experiences of music

In the previous section, we have provided several philosophical observations against Kivy's disanalogy between music and literary texts when repetition is at stake. As we have seen, repetition in literature, similar to music, operates as a key structural and experiential device. Critically, in our opinion, Kivy's disanalogy is also disconfirmed by existing experimental research. In fact, empirical studies indicate that rereading literary texts, as it happens with music, leads to increased appreciation and comprehension.

Let us start by considering the case of music, where Kivy's observations appear to be robustly supported by experimental data. A long tradition of behavioural studies in the 1960s and 1970s, as well as more recent research, has established that the relationship between repetition and pleasure follows an inverted U-shaped curve, or *Wundt curve* (e.g., Jakobovits 1966; Davies 1978; for a critical review, see Margulis 2013). This means that initially, as listeners become familiar with a musical piece, their aesthetic appreciation increases. However, after a certain number of repetitions, appreciation begins to decline, leading to overexposure and, eventually, dislike. One explanation for the initial rise in enjoyment is the *perceptual fluency model* (Bornstein and D'Agostino 1994). According to this model, repeated exposure to a musical piece progressively makes its cognitive processing easier, and the listener tends to mistake this ease of processing for the quality of the music itself, determining an increment of pleasure. Furthermore, as shown in the statistical learning literature (for a review, see Pearce

2018), repeated listening in children and adults can induce implicit internalization of musical patterns and transitions without the need for deliberate analysis.

While this general pattern holds, it has been shown that various factors, including the listening context (Wong et al. 2011), personal history with music (Wong et al. 2012), and psychological characteristics of the individuals (Hunter and Schellenberg 2011), influence the experience of repeated listening (for discussion, see Margulis 2013). For instance, Szpunar and colleagues (2004) demonstrated that the more complex a musical stimulus is, such as orchestral music, the richer the opportunity for enjoyment through repeated exposure. This is because complexity allows listeners to derive more from the music as they become more familiar with its structure. At the neuroscience level, studies using fMRI have also shown that hearing familiar music activates emotion-related brain areas, such as the limbic and paralimbic regions (Pereira et al. 2011), further supporting Kivy's conjecture about the pleasure of repeated listening. Interestingly, recent EEG experiments using the inter-subject correlation technique further illuminate this phenomenon by demonstrating that repeated listening to complex music increases synchronization of neural activity across listeners, particularly in emotion-processing regions like the anterior insula and prefrontal cortex (Madseon et al. 2019). Neural synchronization is, in turn, supposed to be a reliable proxy for aesthetic engagement.

3.2 Repeated experiences of literature

What about repeated experiences of literary texts? Admittedly, empirical research in this area is less developed than in music. Nevertheless, the existing studies seem to directly disconfirm Kivy's hypothesis. For instance, Dixon et al.'s seminal 1998 study aimed to explore the "depth of appreciation" that emerges through rereading (measured on a Likert scale from 0 to 7). The authors' hypothesis was that literary texts, such as Borges' novels, would determine greater appreciation due to rereading as compared to simpler texts like detective stories. Their methodology involved a paradigm in which subjects were asked to read the same text twice (excerpts from Borges' novel *Emma Zunz*) and aesthetically evaluate it after each reading. Results of the study show a significant increase in appreciation for Borges' novel after the second reading.[4]

[4] Note that this does not mean, however, that rereading simpler novels such as detective stories is not correlated to increased aesthetic pleasure. In Dixon et al.'s study, the author used *Death was her Dowry* as control text and found that rereading this novel shows a positive, although modest, increase in evaluation from the first to the second reading (total depth of appreciation = 0.61). For instance, ratings of whether the story was "good literature" increase from 3.87 to 4.13, and recommendations from 3.74 to 4.09, while enjoyment remains stable (4.65 to 4.65). On the other hand, the more complex literary text, i.e., *Emma Zunz*, shows a significantly larger increase in depth of appreciation (= 2.54). This suggests that the relevant difference, therefore, might concern not the presence or absence of

Interestingly, it appears that the critical factor contributing to deeper engagement during the second reading is the increased attention to literary devices like complexity and ambiguity of the narrator. This is suggested by the observation that appreciation rates changed little between the first and second readings when a manipulated version of Borges' novel was used, in which literary effects (i.e., cues to narratorial ambiguity) were removed.

These results were partially supported by a subsequent 2004 study by Hakemulder, which used a similar rereading paradigm. The author used selected excerpts from Rushdie's *The Satanic Verses* (Pilot, Experiment 1), a baroque novel full of complex metaphor, and Nabokov's *The Old Bridge* (64 words, Experiment 1), a shorter and much simpler poem (Experiment 2). In general, as in the 1998 study by Dixon and colleagues, scores on aesthetic appreciation significantly increased after the second reading among individuals with literary training (Experiments 1 and 2), although they remained more or less the same for non-expert readers (Pilot). For *The Old Bridge*, participants were asked a number of questions after both the first and second readings to assess perceived complexity. In this case, readers indicated that "they had enjoyed rereading the poem, that they became aware of aspects in the poem they had not seen the first time, and that they expected to discover even more after reading the text once more" (Hakemulder 2004: 214). Again, increased attention to literary effects during the second reading (i.e., *foregrounding* and deviations from standard language use) appears to be a critical factor: rereading manipulated versions of the same excerpts did not significantly increase appreciation.

In 2018, Kuijpers and Hakemulder conducted three experiments trying to replicate the results of Dixon and colleagues. Participants read either original or manipulated versions of three well-known literary works – i.e., Borges' *Emma Zunz* (Experiment 1), Rushdie's *Midnight's Children* (Experiment 2), Bierce's *An Occurrence at Owl Creek Bridge* (Experiment 3) – and measured their appreciation and perceived comprehension across multiple readings. In the modified versions, complex, ambiguous, or stylistically distinctive features of the narratives were removed, making the texts easier to process and less literary in nature. Self-report measures to test perceived comprehension were used, asking participants to rate how well they understood the literary texts they read. Results confirmed that, contrary to Kivy's supposition, appreciation increased after rereading. However, contrary to previous studies (Dixon et al. 1998; Hakelmuder 2004), this increase was mainly linked to a rise in perceived comprehension, rather than the presence of literary features. Additionally, readers' prior print exposure (reading experience) played a minor role, suggesting that rereading is rewarding for all readers when it increases comprehension, independently of a text's literary complexity.

aesthetic pleasure related to rereading simpler vs more complex literary texts, but the magnitude and nature of the relevant increase.

A final study by Xue and colleagues (2022) examined the effects of rereading on comprehension, fluency, and appreciation using the eye-tracking methodology and testing two of Shakespeare's sonnets, Sonnet 27 ("*Weary with toil*") and Sonnet 66 ("*Tired with all these*"). Eye movements were tracked during both an initial reading and a rereading session to assess real-time cognitive processes. Participants were English speakers with no special literary training. In line with previous studies (Hakelmuder 2004; Kuijpers and Hakemulder 2018), actual comprehension of both sonnets increases significantly, as evidenced by individuals' responses to questionnaires about the main topic of the sonnets. Analysis of objective eye-tracking measures, such as faster total reading times, fewer fixations, and reduced regression times (e.g., the number of fixations of a certain word after the first passage), further confirms that rereading improved subjects' comprehension of the literary works. Surprisingly, rereading did not significantly enhance their appreciation of the sonnets when results were analysed in a combined way. Nevertheless, when the results were analysed separately for each sonnet, there was a slight increase in appreciation of Sonnet 66 after rereading, though not for Sonnet 27.

3.3 Open issues and future research

To summarize, the existing experiments on rereading suggest, contrary to Kivy's disanalogy, that repeated engagements with literary texts can, in a similar way to what happens with music, deepen appreciation and/or uncover new layers of meaning, challenging Kivy's notion that literature lacks the potential for renewed aesthetic pleasure. As we have seen, this seems to be true across literary genres, such as different types of novels (Dixon et al. 1998; Hakelmuder 2004; Kuijpers and Hakemulder 2018) and sonnets (Xue et al. 2022), and not only for experienced readers (but also for readers with no special training (Dixon et al. 1998; Kuijpers and Hakemulder 2018; Xue et al. 2022). Only the 2004 study by Hakemulder found a null effect of rereading on appreciation among non-expert readers, but this result was obtained in the context of a pilot study and using a highly complex and metaphorical text (*The Satanic Verses*), which might necessitate a higher level of literary competence to be fully appreciated. Overall, these findings suggest that the benefits of rereading are not restricted to a niche audience (the "Nabokov-style" reader, in Kivy's words) but extend to a broader readership, suggesting that literature, like music, has a universal capacity to allow for repeatable experiences.

Indeed, we do not want to suggest that Kivy's disanalogy between music and texts has been conclusively falsified by experimental research. Much more research is needed in this field, for instance, by incorporating more ecologically valid stimuli and experimental designs. Traditional experiments on rereading literary texts often use short, isolated texts or excerpts in controlled laboratory settings (e.g., Dixon et al. 1998; Hakelmuder 2004). This methodological procedure may not fully capture the complexities of real-world reading experiences. Using longer texts, such as complete novels, and testing subjects after longer intervals, such as days or weeks, would offer a more authentic view of how rereading unfolds over time and in more natural contexts. In music, this has been done recently by Madson and Schiolde (2018), whose design emphasised ecological validity by allowing participants to repeatedly listen to musical pieces in natural settings, such as at home or during daily activities, over an extended period of time (4 weeks).

Furthermore, potential experiments might directly contrast the willingness to reexperience literary texts and musical pieces that are somehow balanced in cognitive costs, since, as we have seen, reading literature seems to be intrinsically more demanding than listening to music. Future research could also explore how rereading affects appreciation and comprehension across a variety of literary genres, such as formalism, postmodernism, and realism.[5] This establishes a parallel with the domain of music, where some studies have tested the impact of repeated listening on the appreciation of pieces from different genres, like jazz and blues (Orr and Ohlsson 2001). Individual differences also play a crucial role in how rereading affects comprehension and appreciation. Further research might investigate how factors such as prior reading experience, familiarity with literary techniques, cognitive style, or cultural background may shape how a reader engages with a text during subsequent encounters. Through such expanded experimental investigation, we may further support the hypothesis that, like music, literature has a rich potential for repeated aesthetic enjoyment, further challenging Kivy's idea that texts are primarily appreciated for their plot and lose value once the story is known.

As a conclusive observation, it should be noted that an important limitation of current research on rereading is that it is exclusively behavioural. Studies in neuroscience could further substantiate the claim that rereading literary texts and relistening music share cognitive underpinnings. For instance, studies using techniques such as fMRI or EEG might assess whether rereading familiar stories increases activation in emotion-related brain regions, such as the limbic and paralimbic systems, as it happens with music (Pereira et al. 2011; Madseon et al. 2019). Further experiments might investigate inner auditory imagery, a common phenomenon in both reading and music (for discussion, see Barbero and Calzavarini 2024). Researchers might examine whether

[5] For instance, we have seen that, in the 1998 study by Dixon et al., rereading detective stories increases appreciation significantly less than rereading more complex novels such as *Emma Zunz*, suggesting that complexity modulates repeating effects across different literary genres. This, in turn, suggests a potential disanalogy with music, where often simple musical pieces sustain relistening more readily than complex works. Whilst this apparent asymmetry might reflect a difference in modes of engagement rather than a fundamental divergence in repeatability, further empirical work is certainly needed on this issue.

repetition strengthens the vividness and stability of inner auditory imagery in both music and literature, and how individual variability in auditory imagery affects repetition effects across both domains.[6] Finally, experiments could also focus on neural responses to structural repetition, which is a highly underinvestigated phenomenon. For instance, studies could investigate whether repeated encounters with poetic refrains or narrative motifs elicit neural activation patterns similar to those elicited by repeated musical motifs.

4. Conclusion

As we have seen, in *The Performance of Reading* (2006) as well as in other works (2002, 2011), Peter Kivy draws a firm distinction between music and literature, suggesting that the intrinsically repetitive nature of musical pieces makes them better suited for repeated enjoyment than literary texts. Nevertheless, we argue that several strands of theoretical and empirical evidence challenge this disanalogy. At the theoretical level, both music and literature employ repetition as a structural and experiential device that increases appreciation over time. As we have argued, repetition is grounded in cultural traditions, both in the form of recurring motifs in music and in that of refrains and cycles in literature. At the empirical level, extant experiments on rereading reveal that literary texts, much like music, allow for deeper aesthetic and cognitive experiences due to repeated exposure. Although more experimental research is certainly needed in this field, the existing empirical evidence suggests that Kivy's disanalogy underestimates the potential of literary texts to offer repeatable aesthetic experiences similar to those raised by music.

References

Barbero, C., and F. Calzavarini. 2024. "Experiences of silent reading." *Phenomenology and Cognitive Science*. https://doi.org/10.1007/s11097-024-09966-x.

Bornstein, R. F., and P. R. D'Agostino. 1994. "The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect." *Social Cognition* 12 (2): 103–128.

Cappelletti, M., H. Waley-Cohen, B. Butterworth, and M. Kopelman. 2000. "A selective loss of the ability to read and to write music." *Neurocase* 6: 321–332.

Cara, M. A., and G. Gómez Vera. 2016. "Silent reading of music and texts; eye movements and integrative reading mechanisms." *Journal of Eye Movement Research* 9 (7): 1–17.

Clarvoe, J. 2009. "Poetry and repetition." *The Antioch Review* 67 (1): 30–41.

[6] It is known, for instance, that defects in auditory imagery due to hearing limitations or loss affect not only musical but also literary appreciation (for discussion, see Barbero and Calzavarini 2025).

Davies, M. P. 1978. *The psychology of music*. London: Hutchinson.

Di Bona, E. 2019. "Music is not even language-like: Analyzing Kivy's view on music and language." *Argumenta* 8 (2): 71–84.

Dixon, P., M. Bortolussi, L. C. Twilley, and A. Leung. 1993. "Literary processing and interpretation: Towards empirical foundations." *Poetics* 22: 5–33.

Feagin, S. L. 2008. "Critical study: Reading and performing." *The British Journal of Aesthetics* 48 (1): 89–97.

Frank, J. 1968. *The widening gyre*. Bloomington: Indiana University Press.

Gartmann, T. 2021. "Repeatability versus unrepeatability in free improvisation". In A. Bertinetto and M. Ruta (eds.). *The Routledge handbook of philosophy and improvisation in the arts*. Routledge, 392–404.

Hakemulder, J. F. 2004. "Foregrounding and its effect on readers." *Perception and Discourse Processes* 38: 193–218.

Hunter, P. G., and E. G. Schellenberg. 2011. "Interactive effects of personality and frequency of exposure on liking for music." *Personality and Individual Differences* 50 (2): 175–179.

Iser, W. 1984. "The reading process." In J. P. Tompkins (ed.). *Reader-response criticism: From formalism to poststructuralism*. Baltimore: Johns Hopkins University Press, 50–69.

Jairazbhoy, N. A. 1971. *The rāgs of North Indian music: Their structure and evolution*. London: Faber and Faber.

Jakobovits, L. A. 1966. "Studies of fads: I. The 'Hit Parade.'" *Psychological Reports* 18: 443–450.

Kim, G. S., A. Chmiel, and S. Garrido. 2024. "Calming effects of repetition in music for children with sensory sensitivities: Findings from two experimental studies." *The Arts in Psychotherapy* 87: 102115.

Kivy, P. 1993. *The fine art of repetition: Essays in the philosophy of music*. Cambridge: Cambridge University Press.

Kivy, P. 2002. *Introduction to a philosophy of music*. Oxford: Oxford University Press.

Kivy, P. 2006. *The performance of reading: An essay in the philosophy of literature*. Oxford: Wiley-Blackwell.

Kivy, P. 2007. "Music, language, and cognition." In P. Kivy (ed.). *Music, language and cognition: And other essays in the aesthetics of music*. Oxford: Oxford University Press, 214–232.

Kivy, P. 2011. "Paraphrasing poetry (for profit and pleasure)." *The Journal of Aesthetics and Art Criticism* 69 (4): 367–376.

Kuijpers, M. M., and F. Hakemulder. 2018. "Understanding and appreciating literary texts through rereading." *Discourse Processes* 55: 619–641.

Lamarque, P. 2014. *The opacity of narrative*. Lanham: Rowman and Littlefield.

Madison, G., and G. Schiölde. 2017. "Repeated listening increases the liking for music regardless of its complexity: Implications for the appreciation and aesthetics of music." *Frontiers in Neuroscience* 11: 147.

Madsen, J., E. H. Margulis, R. Simchy-Gross, et al. 2019. "Music synchronizes brainwaves across listeners with strong effects of repetition, familiarity, and training." *Scientific Reports* 9: 3576.

Margulis, E. H. 2013. *On repeat: How music plays the mind*. Oxford: Oxford University Press.

Mendez, M. 2001. “Generalized auditory agnosia with spared music recognition in a left-hander: Analysis of a case with a right temporal stroke.” *Cortex* 37: 139–150.

Miller, J. H. 1982. *Fiction and repetition: Seven English novels*. Harvard: Harvard University Press.

Mongelli, V., S. Dehaene, F. Vinckier, I. Peretz, P. Bartolomeo, and L. Cohen. 2017. “Music and words in the visual cortex: The impact of musical expertise.” *Cortex* 86: 260–274.

Nabokov, V. 1980. *Lectures on literature*. New York: Harcourt Brace Jovanovich.

Nanay, B. 2009. “Narrative pictures.” *Journal of Aesthetics and Art Criticism* 67 (1): 119–129.

Nanay, B. 2016. *Aesthetics as philosophy of perception*. Oxford: Oxford University Press.

Nanay, B. 2022. “Why do we reread novels?” *Psychology Today*. https://www.psychologytoday.com

Orr, M. G., and S. Ohlsson. 2001. “The relationship between musical complexity and liking in jazz and bluegrass.” *Psychology of Music* 29: 108–127.

Pereira, C. S., J. Teixeira, P. Figueiredo, J. Xavier, S. L. Castro, and E. Brattico. 2011. “Music and emotions in the brain: Familiarity matters.” *PLoS ONE* 6: e27241.

Ribeiro, A. C. 2007. “Intending to repeat: A definition of poetry.” *Journal of Aesthetics and Art Criticism* 65 (2): 189–201.

Ribeiro, A. C. 2009. “Review: Peter Kivy: The performance of reading: An essay in the philosophy of literature.” *Mind* 118 (469): 186–191.

Ribeiro, A. C. 2009. “Poetry”. In S. Davies, K. J. Higgins, R. Hopkins, R. Stecker, and D. Cooper (eds.). *Blackwell companion to aesthetics*. Oxford: Wiley-Blackwell, 101–104.

Rubery, M. 2016. *The Untold Story of the Talking Book*. Harvard: Harvard University Press.

Szpunar, K. K., E. G. Schellenberg, and P. Pliner. 2004. “Liking and memory for musical stimuli as a function of exposure.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30: 370–381.

Walton, K. 1990. *Mimesis as make-believe: On the foundations of the representational arts*. Harvard: Harvard University Press.

Wong, P. C. M., A. H. D. Chan, A. K. Roy, and E. H. Margulis. 2011. “The bi-musical brain is not two monomusical brains in one: Evidence from musical affective processing.” *Journal of Cognitive Neuroscience* 23: 4082–4093.

Wong, P. C. M., A. H. D. Chan, and E. H. Margulis. 2012. “Effects of mono- and bicultural experiences on auditory perception.” *Annals of the New York Academy of Sciences* 1252: 158–162.

Xue, S., A. M. Jacobs, and J. Lüdtke. 2020. “What is the difference? Rereading Shakespeare’s sonnets—An eye tracking study.” *Frontiers in Psychology* 11: 421.

Yao, B., P. Belin, and C. Scheepers. 2011. “Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex.” *Journal of Cognitive Neuroscience* 23 (10): 3146–3152.

Yu, G. J., and L. F. Ying. 2015. "An analysis of repetitive motifs and their listening duration in selected Western popular songs from 2000 to 2013." *Procedia - Social and Behavioral Sciences* 185: 18–22.

Croatian Journal of Philosophy
Vol. XXVI, No. 77, 2026
https://doi.org/10.52685/cjp.26.77.3
Received: July 1, 2025
Accepted: March 3, 2026

Athenian Democracy: A Debate between its Advocates and Critics

ENES KULENOVIĆ
University of Zagreb, Zagreb, Croatia

The article attempts to reconstruct and offer a critical evaluation of the arguments put forward in debates concerning the nature of democratic rule in ancient Athens between the supporters and the opponents of democracy. By relying on ancient sources, but also insights from contemporary scholarship, the article looks at three main issues discussed in these debates: the issue of efficiency of democratic collective decision-making, the issue of stability of democratic regimes, and the issue of justice of democratic government. Given that the arguments put forward by the critics of democracy are both more voluminous and more philosophically sophisticated, this imbalance in the debate is compensated by the reconstruction of the pro-democratic arguments that appear in the texts of the most prominent critics of democracy in the ancient Greek world, as well as historical accounts offered by modern scholars. In the concluding part of each section, the article offers an assessment of the validity of the arguments of both sides in these debates.

Keywords: Democracy; ancient Athens; Plato; Aristotle; Thucydides.

1. Introduction

Starting with the influential study on the Athenian democracy by Arnold H. M. Jones published in 1953, accepted wisdom among contemporary scholars is that the ancient Greek sources do not provide us with a developed theory of democracy. Jones asserted that "in the abundant literature produced in the greatest democracy in Greece there survives no statement of democratic political theory" (Jones 1953: 1). Three decades later, Moses Finley went a step further, claiming not only that there is no surviving text on democratic theory from that period, but that there never was such a theory: "I do not believe articulated democratic theory ever existed in Athens. There were notions, generalities […] but they do not add up to a systematic theory" (Finley 1985: 49). Subsequently, other scholars have accepted Finley's claim. In her seminal book on the funeral oration in classical Athens, Nicole Loraux talks about "undiscoverable democratic theory" and the "nonexistence of a *written* theory of democracy" (Loraux 1986: 176, 179). Josiah Ober follows suit when arguing that "there was no philosophically articulated democratic theory in fifth- or fourth-century Athens" (Ober 1989: 38) and "there are no surviving texts to explain democratic theory because few such texts ever existed" (Ober 1996: 147). Similarly, Ryan Balot maintains that Athenian "[d]emocrats did not develop a systematic philosophy to support or defend their emphasis on freedom and equality" (Balot 2006: 50).

This insight is often juxtaposed with the fact that political thinkers and writers from the "golden age" of Athens were, for the most part, deeply critical of democracy. As Jones noted: "All the Athenian political philosophers and publicists whose works we possess were in various degrees oligarchic in sympathy" (Jones 1953: 1). Finley agrees: "In antiquity, intellectuals in the overwhelming majority disapproved of popular government" (Finley 1985: 8). Ober confirms this claim: "we have a number of important classical texts critical of democracy but no surviving texts that sympathetically and systematically enunciate the theory on which Athenian democracy was predicated" (Ober 1996: 147). "[A]lmost all known ancient Greek intellectuals", notes Paul Cartledge, were unsympathetic to democracy (Cartledge 2007: 160). The voices of sages critical of democratic rule – those of Thucydides, Socrates, Xenophon, Plato, Isocrates, or Aristotle – have echoed much louder than the voices of the defenders of democracy. However, none of these thinkers, as Kurt Raaflaub pointed out, was an advocate of existing tyrannical or oligarchic regimes (Raaflaub 1983: 517, 522). Philosophers like Plato and Aristotle preferred the rule of the *aristoi* – those who were most intellectually and morally virtuous – and were critical of tyranny and oligarchic rule by the wealthy. Late fifth- and fourth-century critics of democracy, especially after the exclusionary, brutal, and unstable oligarchic governments in 411 and 404 in Athens, were exploring theoretical alternatives to democratic rule while, at the same time, rejecting traditional defence of aristocratic privileges based on birthright and wealth (Ober 1996: 156).

That being said, the conundrum still remains: why, in the birthplace of a democratic experiment that lasted (with only two minor interruptions) throughout the 5th and most of the 4th century (until 322 BC) – the same *polis* and the same period in which the most influential ancient writings in political philosophy were produced – is there no major work of democratic theory? How come the arguments by the critics of democracy are much more prevalent than the arguments offered by the defenders of democracy? Why were there no democratic writers "to defend democracy against what were often violent attacks" (Loraux 1986: 177)? Contemporary scholars of political thought in ancient Athens offer several possible answers to these questions. First, they note that the genre of political theory was developed "by educated elites for consumption by elite audiences of readers or listeners" (Ober 1996: 4). There were not many supporters of popular government among the members of this elite. The most famous written works in that genre were not authored by "democratic leaders" involved in everyday politics of governing the Athenian polis, but by those who "took no part in the affairs of the city" and could indulge "in the activity of writing" (Loraux 1986: 178). Thucydides wrote the history of the Peloponnesian War in exile; Socrates consciously avoided speaking in the Athenian Assembly; both Plato and Isocrates refrained from participating in Athenian politics; and Aristotle was not a citizen of Athens and was therefore unable to participate in its public life.

Second, in contrast to these opponents of democratic rule, advocates of democracy primarily relied on public speeches to put their ideas forward. Public speech, Loraux argues, rather than written work, was the preferred form in which democratic arguments were presented (Loraux 1986: 179). However, the main function of these speeches was to address practical issues of the day. Their goal was not to offer a comprehensive philosophical defence of democracy. Even when democratic values were not merely invoked but elaborated – as in the famous Pericles' Funeral Oration – this did not amount to a fully developed democratic theory. Finally, the dominance of democratic rule in Athens in this period meant that there was no real need for such a theory to be developed. Historically speaking, "democracy arose and was maintained as a practice and as an ideology, it was not the result of a well-heeled, systematic theorist's sketch" (Balot 2006: 50). That means that democrats could rely on a prevalent egalitarian ethos to provide the necessary support for the existing institutions and practices without presenting sophisticated theoretical arguments in writing. "In Athens", Ober argues, "democratic ideology so dominated the political landscape that formal democratic theory was otiose" (Ober 1996: 148). This reveals an interesting point: at the time, the anti-democratic arguments found in works of political theory were not perceived as politically dangerous enough to warrant a philosophically developed written response. Allegiance to democratic principles by the overwhelming majority of Athenian citizens was not shaken by the arguments put forward either by the sceptics or the enemies of democracy. "One defends his faith vigorously", Dennis Peter Maio argued, "when it is under general attack; in fourth-century Athens, the democratic faith was under no such attack" (Maio 1983: 18–19).

Therefore, trying to reconstruct a comprehensive democratic theory from ancient sources seems like a fool's errand. If, as Jones pointed out, there are no surviving sources that provide us with such a theory or, as Finley and later scholars argued, no such sources existed in the first place, any attempt to find a well-developed philosophical defence of democracy from that period leads to an anachronistic trap. However, we can avoid this trap by focusing on a more modest goal of reconstructing the main features of the debate that took place between advocates and opponents of democracy in the Greek world of the 5th and 4th centuries BC. There is an obvious danger of depicting a one-sided debate: not only are the arguments by the critics of democracy much more voluminous, but they are also more sophisticated and developed. Luckily, we can overcome this problem by acknowledging two points. First, adversaries of democracy did not put their arguments forward in a vacuum, but against existing positions and claims defended by their intellectual and political opponents. Even if, as Ober maintains, "[t]he climate of opinion to which the authors of critical political texts were responding was defined less by the reasoned positions of prodemocracy elite intellectuals than by democratic popular ideology and public rhetoric" (Ober 1996: 148), the fact remains that these responses by the critics of the democratic regime offer us insights to both sides of the debate. For a critique to make sense, one must have a certain understanding of the position that is being criticized. Despite one side utilizing complex theoretical concepts and argument, while the other side is relying on "popular ideology and public rhetoric", we can grasp the framework of the debate. The second point is that when reconstructing the arguments of the prodemocratic side of the debate, we are not limited only to the assumed points that critics of democracy argue against or to a few scattered fragments from defenders of popular rule. We are fortunate that in the text of most prominent detractors of democracy – authors like Thucydides, Plato, Aristotle, and even the Old Oligarch – we can find developed prodemocratic arguments. To their credit, instead of constructing a democratic strawman, these thinkers offer us an even-handed account of the arguments put forward by those with whom they disagree regarding the nature of democratic rule. Although these arguments are a far cry from a fully-fledged democratic theory, they are sufficient to provide a deeper insight into the debate on the advantages and disadvantages of democracy in ancient Greece.

The article deals with a debate concerning three aspects of *dēmokratia*: its efficiency, its stability, and its justice. It looks at the arguments from both sides – advocates and opponents of democracy – regarding these three aspects, reconstructing arguments of both sides by relying on both ancient sources and contemporary interpretations of those sources, as well as evaluating the validity of the arguments[1].

[1] The focus of the article, given the origin of the ancient sources, is on Athenian democracy. For other democracies in the ancient Greek world, see: Robinson 2011.

2. Democracy and the question of efficiency

The persistent debate in 5th and 4th century BC in Athens was about democracy's efficiency as a system of collective decision making. The critics were putting forward two mutually enforcing claims: first, that the people – *demos* – lack both expertise and moral character to make responsible and rational political decisions; and second, that *demos* is susceptible to demagogues' manipulation to make wrong decisions. The advocates of democracy, conversely, argued that the people are both capable – through the process of public deliberation – to reach valid decisions and able to make decisions that best reflect the interests of the *polis* as a whole.

A strong claim about the ignorance of *demos* in political matters can be found in one of the first classical texts – Herodotus's *History* – that discusses comparative advantages and disadvantages of different political regimes. In book three, Herodotus describes a debate between Persian leaders that overthrew the brutal and corrupt tyranny of the Magi. This debate, contrary to Herodotus' claim (3.80), is almost certainly fictional. It is very unlikely that the Persians would consider any other options other than monarchical rule for themselves. More likely, the conflicting positions about the benefits and detriments of different regimes presented in *History* reflect debates that Greeks had in the mid-5th century BC about the nature of various forms of government (Winton 2005: 110). First speaker, Otanes, argues for the rule of the many[2] as the best safeguard against the unchecked rule of monarchy that leads to "savage violence" (3.80). Two other speakers, Megabyzus and Darius, reject democracy and, respectively, support oligarchy and monarchy. Megabyzus's concern is that the rule of the people means giving power to "rabble" and "mob" that is "altogether devoid of knowledge" (3.81). This lack of knowledge is two-fold because the rabble, being "untaught", not only lacks the ability to make correct political decisions, but also lacks "natural sense of what is right and fit" (3.81). Therefore, collective decision-making by demos is both epistemically and morally deficient. Megabyzus' claim does not go any further than a mere sentiment, probably widespread among the aristocratic class in Greece at the time, that common people are morally crude and uneducated and, therefore, unable to deliberate responsibly about political matters.

More meat to the bones of this sentiment is given by Thucydides in his *History of the Peloponnesian War* through numerous examples of irresponsible decisions made by the Athenian *demos* during the war against Sparta and its allies[3]. On Thucydides account, Athenian citi-

[2] Otanes talks about *isonomiē* – equality of citizens before the law – rather than *dēmokratia*. For a more detailed discussion on Herodotes's use of term *isonomiē*, see: Cartledge 2007: 158–159.

[3] For a more detailed account of Thucydides' views on democracy, see: Raaflaub 2006; Mitchell 2015, ch. 5.

zens in the Assembly, often poorly informed and ruled by emotions, made repeated bad policy decisions that, in the end, resulted in Athens' defeat in the war[4]. The numerous examples of collective incompetence and recklessness he offers suggest that *demos*, unless guided by a capable political leader, has a tendency to make uninformed and politically questionable decisions, as well as a propensity for refusing to take responsibility for those decisions. In the second book, Thucydides praises Pericles as a true statesman who led the people "rather than they who led him" (2.65), arguing that democracy only works, especially during a crisis, when governed by such great men. Paradoxically, democracy functions only when it is not a true democracy, i.e., when "what was nominally a democracy, power was really in the hands of the first citizen" (2.65).

Plato's criticism of democratic decision-making is a part of his larger philosophical project that relies on "a formal distinction between mere opinion (*doxa*) and actual knowledge (*epistēmē*)" (Ober 1996: 158). Plato's dialogues also offer examples of the majority making wrong and morally dubious judgements, such as Socrates' retelling in the *Apology* of the Assembly's unlawful decision to condemn their own naval generals to death after the Battle of Arginusae. Socrates, the only member of the Council to vote against this illegal verdict, faced arrest by the orators supported by the people whose "shouts were egging them on" (32b-c). For Plato, as revealed in the *Seventh Letter*, the most disturbing case of wrongful decision-making by the majority was when, after democracy was restored in Athens and general amnesty was declared, the jury of 500 nevertheless "condemned and put to death" (325b-c) Socrates, his great teacher, friend, and "the justest man of that time" (324e). For Plato, these examples are indicative of a deeper problem of political and legal decision-making by the *demos*: the majority simply lacks the necessary knowledge, both on technical and moral issues, to reach responsible and sound judgments. In *Protagoras,* Plato's Socrates points out the irony that when it comes to technical matters – "building projects" or "construction of ships" – people rely on the experts to

[4] Thucydides' examples of reckless decision made by Athenian Assembly are numerous: instigating the conflict with Sparta's main ally Corinth by supporting Epidamnus (1.24-1.55); denying their responsibility and blaming Pericles for the decision to start the war against Peloponnesians (2.53.2-3); disregarding the neutrality of Melos and brutally conquering this island, executing all men and selling all the women and children into slavery (5.116). In their desire to conquer Syracuse and control Sicily, although "for the most part ignorant of the size of the island and of the numbers of its inhabitants" (6.1.) they voted for the Sicilian Expedition, which resulted in Athens' catastrophic defeat. Even those who were opposed to this expedition, did not vote against it because they feared to be seen as unpatriotic (6.24.). When the news of this defeat reached Athens, they first refused to believe it and later denied their responsibility in supporting this expedition by blaming orators that encouraged the Assembly to send the fleet to Sicily (8.1). Finally, in 411 it was the majority of citizens that voted to abolish democracy and allow Pisander and his supporters to establish oligarchy in Athens (8.63).

advise them and will only take into account the opinions of "craftsman" as valid (319b-c). However, "when it is a matter of deliberating on city management", the citizens in the Assembly will listen to anybody: "carpenter, blacksmith, shoemaker, merchant, ship-captain, rich man, poor man, well-born, low-born" (319d). Democracy, according to Plato, rests on an epistemic error by assuming that political decision-making, unlike most other forms of decision-making, does not require expertise and knowledge.

Interestingly, we can rely on almost the same ancient sources for counterarguments in the debate about the efficiency of democratic decision-making. Thucydides offers us two notable instances of praise for the ability of the *demos* to make responsible decisions. In the first year of the Peloponnesian War, Pericles, in his famous Funeral Oration, claims that in Athenian democracy, citizens show "care both of their own and the public affairs" and have "sufficient knowledge of state matters even in those that labour with their own hands" (2.40). He emphasizes how fellow Athenians "excel others" by "daring to undertake as much as any and yet examining what we undertake" and using public deliberation to "weigh what we undertake and apprehend it perfectly in our minds" (2.40). Freedom to discuss and make decisions collectively plays a crucial role in Athenian greatness. As already noted by Herodotus (5.78), the success of Athens as, economically, militarily, and culturally, the most prominent *polis* in the Greek world in the mid-5th century BC offered a best defence of its democratic regime. If democracy inevitably leads to blunders and bad decisions by the *demos*, it would be hard to explain Athens' accomplishments[5]. The second instance happens much later in the war in Syracuse, during a debate on the dangers of the Athenian invasion of Sicily. Athenagoras, the Syracusan "chief magistrate of the people" (6.35), counters the oligarchic claim that democracy is neither just nor well-governed by pointing out that oligarchy represents only the interest of the wealthy, while democracy, by allowing everyone to participate in the decision-making process, represents the *polis* as a whole. He concedes that "rich are indeed fittest to keep the treasure" and that "the wise are the best counsellors", but that "multitude, upon hearing" is "the best judge" (6.39).

One of the most passionate defences of the democratic ethos comes from Protagoras in Plato's dialogue of the same name[6]. As a response to Socrates' claim that the people lack the necessary knowledge to

[5] As Raaflaub points out: "Success in leading Athens to greatness was one of the strongest arguments in support of democracy and difficult to contest as long as success lasted. The setbacks in the latter part of the Peloponnesian War deprived democracy of this crucial source of legitimation and made it vulnerable" (Raaflaub 2006: 211).

[6] For Protagoras as "the first democratic political theorist in the history of the world", see: Farrar 1988: 77. For more detailed discussion on Protagoras' democratic arguments, see also: Finley 1985: 28–29; Ober 1996: 126, 129, 130; Balot 2006: 75–78; Mitchell 2015: 64.

make sound political decisions, Protagoras argues that all people have a sense of justice (*dikê*) and moderation (*aidôs*) and, therefore, the ability to develop political excellence (*politikê aretê*). His argument is presented as a myth in which humans, to protect themselves from wild beasts, founded early cities. Because they lacked "the art of politics", they started fighting each other, and cities were destroyed or abandoned (322b-c). Zeus, being afraid that the human race would go instinct, ordered Hermes to "bring justice and a sense of shame to humans, so that there would be order within cities and bonds of friendship to unite them" (322c). To ensure stability, Zeus told Hermes to distribute these traits equally to all humans. For Protagoras, Socrates is wrong to equate moral and political judgment (which is "shared by all") with technical expertise such as architecture or shipbuilding (which is in the hands of experts). Without "civic virtue" being "shared by all […] there wouldn't be any cities" (323a). The ability to make responsible political decisions – civic virtue – is not reserved exclusively for experts, educated aristocrats, or philosophers. That is why, Protagoras points out to Socrates, there is a "good reason that your fellow citizens accept a blacksmith's or a cobbler's advice in political affairs" (324c-d).

Aristotle offers a more technical defence of collective decision-making in democracy. In Book III of *Politics,* he presents his "summation argument": although most individual members of the multitude have less knowledge and understanding than an educated person, when deliberating together and when "regarded not individually, but collectively" (1281b2), their decision can be equally sound or possibly even better. The many, Aristotle maintains, "individually they may be worse judges than those who have special knowledge, as a body they are as good or better" (1282a16-17). There are two reasons for this. First, as "each individual among the many has a share of excellence and practical wisdom" (1281b4), when these shares are combined, they offer both better insight and a diverse perspective on the issues being discussed and decided upon. This kind of multiple perspective – where "some understand one part, and some another, and among them they understand the whole" (1281b9-10) – leads to better decision-making. Second, decisions by the many are more likely to reflect the interests of the whole *polis* rather than just the interests of the rich and educated aristocrats, even if these aristocrats have more expertise in political matters. As the knowledge of what is a good house is not limited only to an architect, but also to the occupant of the house, and as the pilot of the ship can be a better judge of the rudder than the carpenter who made it (1282b20-23), so the majority of citizens can better decide what is in the interest of the *polis* than a few experts. Irrespective of his "doctrine of the wisdom of the multitude"[7], Aristotle does hold an elitist

[7] On interpreting Aristotle's "doctrine of the wisdom of the multitude" as being about multiple perspectives and knowledge, see: Waldron 1996. For different interpretation, where collective decision-making is primarily about developing virtue (*aretē*), see: Cammack 2013.

view that, individually, members of the *hoi polloi* are more corruptible and prone to errors in political judgment and, therefore, "[t]here is still a danger in allowing them to share the great offices of state" (1281b25-26). However, this view is compatible with political practices in democratic Athens through election of ten generals (*strategoi*), based mainly on their strategic expertise; the advantage of speakers in front of the Assembly who were educated in rhetoric; and the Assembly's reliance on the advice of experts when it came to technical issues such as wall-building, ship-building, or finances (Mitchell 2015: 55, 238–239; Hornblower 1993: 13). Later, in the 4th century, the *Theorie* Fund was established as a professional class of elected experts who dealt with state finances (Ober 1989: 102).

The second line of criticism relating to limited epistemic abilities of the people refers to the danger of demagoguery in democracy. On this account, the common men that make up the *demos* are often susceptible to manipulation by power-hungry demagogues[8]. Rhetorical manipulation of the people comes in two guises: either a) by demagogues strengthening their own political standing and support by encouraging irrational impulses and prejudices of the people, or b) through promoting policies that benefit demagogues' narrow self-interest by masking them as policies that would benefit the *polis* as a whole. Both pandering to and manipulating *demos* by political leaders and *rhetors* lusting for personal power are recurring themes in ancient sources. Herodotus tells us how Pisistratus deceived the people of Athens twice, first to gain power as a tyrant and then, after his exile, to reestablish himself (1.59-60). Or how Aristagoras of Miletus, after being rejected by the Spartans, managed to convince the Athenians to grant him twenty ships for an unnecessary naval expedition that resulted in provoking Persian king Darius, leading to Herodotus' conclusion that "it seems indeed to be easier to deceive a multitude than one man" (5.97). Thucydides gives an account of both manipulation and pandering. Unimpressed with the political leadership of Athens after Pericles' death, he points out that "his successors, who were more on a level with each other and each of whom aimed at occupying the first place, adopted methods of demagogy which resulted in their losing control over the actual conduct of affairs. Such a policy, in a great city with an empire to govern, naturally led to a number of mistakes" (2.65). Two most prominent examples of demagoguery by Pericles' successors in Thucydides' narrative are those of vulgar and brutal Cleon and charismatic and duplicitous Alcibiades. Cleon "a popular figure […] who had the greatest influence with the masses" convinces Athenians to reject Sparta's peace offer (4.21). Later, when a crucial Athenian blockade of Pylos crumbles and regret for rejecting the truce spreads among the

[8] Historically, *demagogoi* – leaders of the people – was not a pejorative term, referring to orators that promoted interest of the people. However, in late 5th and throughout 4th century this term acquired a negative meaning that it has today (see: Finley 1985: 38–75; Ober 1989: 16–17; Mitchell 2015: 149–150).

people, Cleon, "becoming unpopular", falsely claims that reports from Pylos are untrue. He manages to persuade his fellow citizens to launch the Pylian campaign (4.27), which, against all odds and despite his military inexperience, proves successful (4.41). A decade later, in 415, Athenians were less lucky with a disastrous Sicilian Expedition they decided to launch at the behest of Alcibiades, who, although an able general and a great public speaker, was motivated first and foremost by his personal "wealth and honour" (6.15).

In the *Republic,* Plato deals with both sides of the coin of demagoguery – manipulating the multitude and pandering to the multitude. In the famous metaphor of the ship, the shipowner, i.e., the people, who is "bigger and stronger than everyone else on board" is also "hard of hearing, a bit-short-sighted, and his knowledge of seafaring is equally deficient" (488a-b). The sailors, i.e., demagogues, although they have no knowledge of the art of navigation, are fighting each other to become captain and control the ship by "crowding around the shipowner, begging him and doing everything possible to get him to turn over the rudder over to them" (488b-c). They manage to "rule the ship" by stupefying "their noble shipowner with drugs, wine, or in some other way" (488c). Later in the same book, Plato talks about sophists using their oratory skills to appease "the moods and appetites of a huge, strong beast", the people (493b). This complex relationship that exists in democracy between playing up to people's whims and deceiving them is also explored in Plato's critique of rhetoric, *Gorgias*. In this dialogue, Socrates claims that rhetoricians have to rely on the ignorance of their public to make their arguments persuasive (459a). The aim of rhetoric is to persuade an audience to hold a certain conviction on the subject, not to reach a true knowledge of the subject. The teacher of oratory, Gorgias, insists on the importance of his craft by offering an example of going with his brother, who is a doctor, on calls to sick people and managing to convince patients, who are unconvinced by his brother's arguments, to take the medicine or allow a doctor to operate on them (456b). However, this example only plays into Socrates' hands by confirming his point that experts like doctors have a necessary knowledge that orators do not, and that they only need orators' skill of persuasion when faced with "those who don't have knowledge" (459a). The same is true in the democratic assembly. Later in the text, Plato departs from Thucydides' positive portrayal of Pericles by describing him as one of the demagogues who feasted the people "lavishly with what they had an appetite for" – walls, dockyard, harbours, and tribute payments – instead of "justice and self-control", making the city "swollen and festering" (518e-519b). In *Politics*, when discussing different forms of democracy, Aristotle points out that when the multitude has supreme power and rules by decrees without being bound by laws, a demagogue, as the flatterer of the people, will "have great power" (1292b21). In the *Constitution of Athens,* Aristotle offers historical examples of "the people being led astray" by orators when condemning to death their own generals after the Battle of Arginusae and, later, refusing a peace offer from Lacedaemonians (34).

The ancient sources do not provide a direct rebuttal of the charge that *demos* is susceptible to rhetorical manipulation. However, there is evidence that Athenians were well aware of the dangers of being indulged or manipulated by skilful public speakers. In Aristophanes' comic play *Knights,* produced in 424, Cleon is ridiculed as a crude populist, manipulating the people for his own benefit, using trumped-up accusations against his political opponents, using flattery of the *hoi polloi* to gain support, and lining his own pockets from the public treasury. When, at the end of the play, the chorus accuses the character Demos of being naïve and ignorant, Demos reveals that it is only playing being dumb, while keeping a watchful eye on those who think they are cheating him, ready to strike them down when necessary (1121-50). Thucydides, in several instances, clearly points out how suspicious citizens of Athens could be of their political leaders. After defeating Mytilene, the initial brutal decision by the Athenians to kill all the men and sell the women and children into slavery, a policy advocated by Cleon, is reversed by Diodotus' appeal to moderation (3.42). When Cleon advocates for the Pylian campaign and tries to convince the Assembly to give command of the campaign to experienced military commander Nicias, the citizens decide to support Nicias' advice and make Cleon the leader of the campaign (4.28). Although Athenians found out about Alcibiades' treachery to Sparta only after his death, Thucydides points out that many citizens were aware of his personal ambitions and tyrannical tendencies (6.15). Conscious of the dangers of demagoguery and the use of public support for personal gain, Athenian democracy introduced mechanisms such as ostracism and *graphe paranomon* to try to address these dangers (see: Finley 1985: 26–27; Ober 1989: 95–96; Mitchell 2015: 76). Ostracism, an opportunity to exile any citizen for 10 years if they were perceived too influential by the public, was introduced specifically to address the issue of powerful individuals that could subvert democracy. During the Peloponnesian War, Athens introduced the *graphe paranomon,* which enabled the legal prosecution of anyone who proposed a policy that was contrary to the law and against the public interest, even if that policy was accepted by the Assembly. Both mechanisms illustrate awareness of democratic forces in Athens about possible shortfalls of collective decision-making and their ability to self-regulate and self-correct those shortfalls. The critics pointing out the negative influence of demagogues in democracy were not wrong, but they might have underestimated the capability of democracies to temper such influences. If demagoguery finds fertile ground in democracy, it does not mean it is also inevitable.

3. Democracy and the question of stability

The critics of democracy were sceptical about the stability of a democratic regime, worrying that it is prone to internal power struggles – *stasis*[9] – that could escalate into civil war. Additionally, there was a concern that democracy's inherent factionalism and volatility encourage people to seek stability by relinquishing the state's power in the hands of the established elite or one person, resulting in oligarchy or tyranny. On the other side of the debate on democracy's stability, we find arguments – even among some opponents of popular rule – claiming that democracy is not inherently more unstable than other forms of government or that, as the historical record shows, it is even more politically stable than oligarchies or kingships.

Already in Herodotus' aforementioned Persian Debate, we can find elements of the argument that democracy leads to *stasis* and the argument that the end result of its instability is the establishment of a monarchy. Darius, as an advocate of monarchical rule, claims that oligarchies lead to power struggles among prominent men, and that such strife then often ends "in bloodshed". Democracies, conversely, lead "to close friendships" among those who have carried out "villainies" together, i.e., factionalism of competing "evil-doers". The solution to both oligarchy and democracy is monarchy as a guarantor of both stability and justice (3.82). In his narrative of the Peloponnesian War, Thucydides offers a sobering account of how, in turbulent times, both in oligarchies and democracies, moderation gives way to self-interest, conflict, and violence. The most dramatic depiction of this insight is that of civil strife in Corcyra (3.69-84) between members of the pro-Athenian democratic party and pro-Corinthian members of the oligarchic party. Internal conflict among these two parties, exacerbated by the external threat of nearby Corinthian and Athenian fleets, resulted "in massacre […] of their own citizens" whom the more dominant democratic faction "considered to be their enemies" (3.81). This type of "violent fanaticism" caused by "love of power, operating through greed and through personal ambition" (3.82) spread to other Greek city-states, societies "divided into two ideologically hostile camps" (3.83) professing to "serve public interest" – either "political equality of the masses" or "sound government of the aristocracy" – while, in reality, "they were seeking to win prizes for themselves" (3.82). Thucydides shows a devastating effect that *stasis* has on both democracies and oligarchies. However, when focusing on Athenian democracy, Thucydides argues that without proper political leadership, without someone like Pericles at the helm, democracy's internal divisions contributed in large part to Athenian defeat in the Peloponnesian War. As Josiah Ober concludes, "[*d*]*ëmokratia*, as understood by Thucydides, was the unstable rule of mob and demagogues ", that is, "inherently unstable form of political

[9] For linguistic origin of the word *stasis*, see: Finley 1985: 44–45.

organization, one that, more or less predictably as a result of its own internal contradictions, pulled itself apart under the stress of a long struggle with a hostile nondemocratic power" (Ober 1989: 130–131).

The relationship between democracy and *stasis* was also explored by Aristotle in *Politics*. In Book V, when discussing political revolution, he offers several examples of democracies – in Thebes, Megara, Syracuse, and Rhodes – that collapsed due to "disorder and anarchy" (1302b30-34). When later discussing the reasons for this, he concluded that "revolutions in democracies are generally caused by the intemperance of demagogues" (1304b21-22), who, in "order to curry favor with the people, wrong the notables and so force them to combine" (1305a3-4), resulting in the overthrow of the democratic regime. In a desire to increase personal power and wealth, demagogues stir people against the rich, confiscate their property, and bring false accusations in courts against them, forcing them to unite and take over political control. Aristotle illustrates this argument by talking about oligarchic takeovers in Cos, Rhodes, Heraclea, Megara, and Cyme (1304b25-39). *Politics* also offers insight into the relationship between *demos* and tyranny[10]. When discussing types of monarchies, Aristotle notes that elective tyranny – *aisumnētēs* – arises from popular support (1285a29-b2). For Aristotle, the main reason, at least in archaic times, that the *demos* decided to support tyrants – like Pittacus in Mytilene – was to curtail aristocratic abuse and ensure stability.

His teacher Plato makes an explicit argument about the close connection between democracy and tyranny. In Book VIII of the *Republic*, when talking about different types of constitutions and their specific weaknesses, he explains how tyranny comes into being as a consequence of democracy's instability. The irony, for Plato, is that "democracy's insatiable desire for what it defines as the good" – freedom – is "also what destroys it" (562b). Its obsession with avoiding "having any master at all" leads to ignoring laws and to "general permissiveness" that "eventually enslaves democracy" (563e). Plato elaborates this process in the following way: the idle and extravagant men dominate the city by rallying otherwise passive people against the rich, distributing the wealth among the people, but "keeping the greater part for themselves" (565a). When the wealthy complain, they are accused of plotting against the people and trying to establish an oligarchy. This turns into a self-fulfilling prophecy, as the conflict between the people and the wealthy escalates. Class conflict is inevitable in a democracy. To resolve this conflict, the *demos* elects "one man as their special champion" (565c) who soon turns into a tyrant. To win support and keep himself in power, a tyrant first distributes land and wealth, but then proclaims war, raises war taxes, gets rid of any critics or perceived political opponents, and surrounds himself with a private army of bodyguards. By

[10] For origin and the use of term *tyrannos* in archaic and ancient Greece, see: Salmon 1997; Parker 1998.

spilling "kindred blood" and using violence against the people, he turns himself into a wolf (565d-e). In its endless desire for freedom, democracy puts itself into slavery, i.e., its inherent instability leads to tyranny.

Were these assertions about the unstable nature of popular rule justified? The main line of argument against the claim that democracy is inherently unstable is historical, rather than philosophical. True, there were periods in the history of Athens or in other democratic city-states when internal conflict led to bloody civil strife or to tyranny supported by the *demos*. Apart from cited historical examples offered by Thucydides and Aristotle, the obvious case is that of Pisistratus, who came to power and established his tyranny in Athens in 546 BC with strong public support[11]. However, as contemporary scholars point out, when it comes to Athenian democracy in the mid-5th century and especially 4th century BC, the democratic government proved to be quite stable.

Finley argues that for almost two hundred years of democratic rule, Athens achieved stability, political, and military power, as well as economic and cultural prosperity (Finley 1985: 23). After the Peloponnesian War ended and throughout most of the 4th century BC, there were no major internal conflicts between the mass and the elite, no looming danger of tyranny or disregard for the law, no policy of extensive land or wealth distribution, and no support for extreme demagogues or attempts at an oligarchic coup (Ober 1989: 55; Mitchell 2015: 255). Ober concludes that this historical record shows that the "stability of democracy" was not "the result of inertia or a historical fluke" (Ober 1989: 18).

Ancient sources confirm this conclusion. In the text *The Constitution of the Athenians,* written around 420 BC by an anonymous author hostile to democracy, known only as the Old Oligarch[12], popular rule is decried as unjust and unable to achieve good government (*eunomia*). However, this anonymous author also acknowledges that Athenian democratic rule is efficient and stable. He claims that the conflict of interest between the rich and the poor, and between the aristocrats and the *hoi polloi,* is unavoidable, but that democratic Athens has proved very successful in neutralizing this conflict and preserving their constitution (1.1). They achieved this by keeping "the generalships or calvary commands" in "the hands of the most influential men" (1.3), allowing everyone – both highborn and lowborn – to participate in the process of political decision-making (1.6), not allowing great economic inequality (1.4), relying on naval power, which is mainly in the hands of the people and not high-born elites (1.2), and by promoting the interests of the majority (3.10) without prosecuting or disfranchising the rich (3.12).

[11] For more details on Pisistratus' tyranny, see: Ober 1989: 65–67; Mitchell 2015: 32–34.

[12] On Old Oligarch, also known as Pseudo-Xenophon, see: Ober 1998: 14–26; Balot 2006: 92–97.

Democracy is designed for the benefit of the poor and uneducated, not for wealthy, high-born "good men," and is therefore, from the perspective of the author of the *Constitution of the Athenians*, unjust. Nonetheless, the people "have preserved the democracy well" (3.1). For the Old Oligarch, justice and stability do not necessarily go hand in hand: Athenian democracy is an example of an unjust yet stable government.

Aristotle, in both *Politics* and *Constitution of Athens*, describes a process of institutional innovations in democratic Athens that were primarily designed to safeguard stability by addressing the problems of *stasis* and tyrannical tendencies of ambitious individuals. These innovations included the loss of civil rights of any person trying to establish tyranny, ostracisms, *graphe paranomon*, magistrates' obligation to rely only on written law, and establishing a more moderate form of democratic rule from 403 BC that put the rule of law above the will of the *demos*[13]. In conclusion, regarding the issue of democracy's stability, we can draw two inferences. First, although there are numerous historical examples of democratic *polises* falling into *stasis* or tyranny, democracy as a political regime was not inherently unstable or, as Plato argued, destined to end in tyrannical rule. Plato's insight that tyranny follows in the footsteps of democracy is valid if read as a warning, but questionable if it implies inherent instability of democratic rule that inevitably leads to tyranny. As documented by Thucydides and Aristotle, oligarchies were equally or even more likely to succumb to civil unrest or support for tyrannical solutions, especially in the context of all-out-war in the Greek world.[14] Second, the introduction of the abovementioned institutional mechanisms in Athens suggests that democrats were aware of the dangers of popular tyranny and civil strife and were ready to use legal checks and institutional innovations to curtail these dangers.

4. Democracy and the question of justice

The last debate I want to look at deals with the justice or injustice of democratic rule. The exponents of *demokratia* – such as the leader of the democratic party in Syracuse, Athenagoras – argued that democracy is the most just form of rule because "members of the same State ought, in justice, to enjoy the same rights" and because this type of rule accounts for the interests of all citizens. "What is meant by the *demos*, or people, is the whole State", Thucydides reports Athenagoras' claim, "whereas an oligarchy is only a section of the State" (6.38-39). In contrast, the critics saw democracy as an unjust system of rule because of its blind egalitarianism, because it promotes the interests of

[13] For more detailed account of institutional reforms in Athens, see: Aristotle's *Constitution of* Athens, 16 (on law against tyranny), 22 (on ostracism), 42-69 (on legal reforms); Finley 1985: 26–27; Ober 1989: 73–75; Mitchell 2015: 201–208.

[14] Winters notes that "the key source of *stasis* during fifth and fourth centuries was actually from within oligarchy" due to "ambitions and confidence of predatory oligarchs" (Winters 2011: 86).

the poor, crude, and uneducated majority at the expense of the wealthy aristocratic minority, and because it leads to expansionist and imperial foreign policy.

The Old Oligarch complains that by choosing a democratic constitution, Athens chose "to let the worst people be better off than the good" (1.1). His argument throughout the text relies on an oligarchic understanding of the natural superiority of the few and the rejection of Protagoras' egalitarian idea that civic virtue is shared by all. However, this elite ideology of claiming special status based on "high birth" was successfully countered by democratic ideology that depicted all citizens of Athens, irrespective of their wealth or social rank, as *autochthones* descendants of "earth-born" inhabitants of Attica (Ober 1996: 26–27). The claim about the natural superiority of the members of the aristocratic class probably lost most of its argumentative persuasiveness in Athens after the oligarchic coups of the Four Hundred in 411 and the Thirty Tyrants in 404, which revealed their incompetence, greed, brutality, and complete disregard for their fellow citizens (Raaflaub 1983: 535). Unsurprisingly, the philosophers in the 4th century, when talking about the injustice of democratic rule, did not rely on arguments about the natural hierarchy of birth and wealth. When Plato's Socrates in the *Republic* talks about how democracy promotes "equality to both equals and unequals alike" (558c), his concern is that it unjustly neglects true merit. For Plato, merit is not based on traditional oligarchic indicators such as wealth or lineage in a prominent family, but on objective knowledge of what is true and just, as well as a moral character to act in accordance with that knowledge. The members of the ruling class in the ideal city – *kallipolis* – depicted in the *Republic* are chosen through a rigorous meritocratic system of education that is designed to reward individual talents and nullify the privileges of wealth and high-birth. These rulers are not a traditional elite, but an epistemic one, chosen to lead the city because of their knowledge of what is good for the whole *polis*. Their ascetic existence stands in clear contrast to the oligarchic desire for material wealth and also to Plato's description of the democratic man who, in his ignorance, "declares that all pleasures are equal and must be valued equally" (561c).

We can find a similar rejection of traditional oligarchic privilege in the third book of Aristotle's *Politics,* when he argues that "advantages of wealth and birth" do not in any way guarantee excellence in matters of politics. He uses the example of flute-players in which one's wealth or birth in a prominent family does not "contribute to excellence in flute-playing" (1283a1). As a person's merit in music is determined by his excellence in playing the instrument, "the rival claims of candidates for the office can only be based on the possession of elements which enter into the composition of the state", i.e., excellence in political prudence[15].

[15] It is worth noting that Aristotle allows that "well-born" and "rich" might have a good reason to claim office based on the mere fact that they are often more educated

Aristotle's insistence on moral virtue and prudence as the main criterion of good citizenship leads him to criticize not only traditional hierarchies but also democratic egalitarianism. "The basis of democratic state is liberty", argues Aristotle in Book VI of *Politics*, and that liberty manifests itself in two ways (1317a40). First, there is the principle that "man should live as he likes" (1317b11). Second is "for all to rule and be ruled in turn" (1317b1-2). This entails "the application of numerical not proportional equality", where the will of the majority is "the end and the just" (1317b4-6). Neither of these principles of democratic justice caters to the ideal of meritocratic excellence that Aristotle advocates.

Is the charge that democratic egalitarianism leaves no room for merit justified? Looking at both democratic rhetoric and practices in Athens suggests otherwise. In Thucydides' retelling of Pericles' famous funeral oration, Pericles affirms the egalitarian nature of the Athenian constitution, where "power is in the hands not of minority but of the whole people" (2.37). However, he also emphasizes the importance of democratic merit, which is reflected in the "service to the state" (2.37) rather than just personal excellence. Therefore, the citizen who is poor can contribute to the *polis* and be meritorious as much as a wealthy citizen[16]. Democrats did not dismiss the value of merit as undemocratic but redefined it in more egalitarian terms. They did, though, leave enough room for individual excellence as pointed out in Pericles' claim that when it comes to "putting one person before another in position of public responsibility" what is taken into account is "actual ability which the man possesses" (2.37). Election to public offices on the basis of merit, in contrast to election by lot, was considered an aristocratic element in Athenian democracy. Nevertheless, it was an important part of the democratic constitution, which is demonstrated by the fact that offices that required military or financial skills were elective and, therefore, most often filled by members of the upper classes, such as Pericles (Mitchell 2015: 250). Similarly, although all citizens, regardless of class, had the right to address the assembly, it was mainly those who were educated in the skills of oratory who did so (Hansen 1991: 306-20; Balot 2006: 82-84). In practice, Athens combined democratic egalitarianism with aristocratic excellence.

The most prominent argument about the injustice of democratic rule was based on a claim that this type of rule promotes the interests of the majority at the expense of the minority. Old Oligarch complains that the majority is "not so much concerned with justice as with their own advantage" (1.13) and that the best element in society – the

and, as tax-payers, contribute more to the state (1283a16). He does, however, point out that the state can exist without virtues of "justice and valour", but "not well" (1283b21).

[16] Rich citizens can help the state by paying more taxes, but poor citizens contribute to the assembly, as jurors, as soldiers in war, and as rowers of trireme. "Pericles' funeral oration", points out Monoson, "equates elaborate acts of munificence with modest acts of participation" (Monoson 1998: 501).

noble, educated, and the rich – are extorted through funding public feasts, theatre, and athletic contests in which the people then participate. He holds a cynical view of politics, rejecting the idea of the common good and claiming that in oligarchies the rich minority rules, first and foremost, for its own benefit. In the same way, in democracies, the poor majority rules in its own interest. The rich want to protect their wealth, while the poor want to protect their freedom, yet they both want to rule. The crucial difference, from the Old Oligarch's perspective, is that the educated and wealthy, unlike the crass and the poor, can ensure good government (*eunomia*). Similarly cynical view of the nature of class interest in politics can be found in Thucydides when he tells us about Spartan general Brasidas speaking in front of Acanthians, assuring them that he did not come to pick sides between the democratic and oligarchic factions or to "enslave many to the few or the few to the many" (4.86). Here, "enslaved" is a figure of speech: neither were the majority enslaved in oligarchy, nor were the rich few enslaved in democracy. However, oligarchies excluded the majority of citizens from the decision-making process, while democracies were inclusive of all citizens. The way that the noble and the rich could claim they were enslaved in democracy is that they were "overpowered" by "the will of the whole polis, which included their wills without it being their will" (Raaflaub 1983: 525).

This is the reason why Aristotle, when classifying different types of political regimes in the third book of *Politics*, talks about democracy as one of the "despotic" and "defective and perverted forms" of rule (1279a20-21). Like tyranny or oligarchy, democracy promotes the interests of the ruling class, not the common good. In the case of democracy, that means the interests of the "needy" and poor majority at the expense of everyone else (1979b9). Later, in book four, Aristotle goes into more detail by describing four stages of democracy, offering us a narrative of decline from constitutional government as a more moderate type of democracy based on the rule of law to an extreme form of democracy "in which the people are superior even over the laws" (1298b15-16)[17]. However, unlike the Old Oligarch and Brasidas, Aristotle assumes that there are forms of government that benefit both the rulers and the ruled: a moderate form of democracy – constitutional government (*politeia*) – is one such example where laws are designed to benefit all citizens.

The main argument against democracy as a system that promotes the interests of the poor at the expense of the wealthy focused on the perceived injustice of property and wealth distribution, as well as on the financial burden placed on rich citizens to contribute to maintaining public goods. Advocates of traditional aristocratic privilege, such as The Old Oligarch, bemoaned the fact that the rich citizens' overwhelm-

[17] For criticism of Aristotle's "unrelenting theory-based aversion to radical democracy", see: Mitchell 2015: 207–208.

ing contribution to the state –through paying for liturgies and taxes, fighting in the cavalry, and providing expensive war equipment – was not reflected in their diminishing political influence (Raaflaub 1983; Ober 1996). In 462, institutional reforms lowered the property qualification for office holding and introduced salaried government service. Two decades later, pay for jurors in public courts was instituted. In the aftermath of the Peloponnesian War and throughout the 4th century, Athens could no longer rely on profits from silver mines in its colonies or revenue from allies in the Delian League, so more financial burden fell on rich citizens – through voluntary contributions and taxes, as well as securing pay for attending the Assembly – to keep the state solvent and its democracy functioning (Ober 1989: 30, 99, 133). From the perspective of the wealthy, this was clear proof of the unjust nature of democracy as a system that maintains itself by extorting a more capable minority for the benefit of a less successful majority. However, for the advocates of democratic rule, programs of wealth distribution served two main purposes: 1) use of private wealth to subsidise the poorest citizens and allow them to participate in politics, and 2) to control the affluent elite from using their wealth to make themselves too influential or immune to law (Ober 1996: 184)[18].

Were redistributive policies in democratic Athens unjust? The answer to that question depends on one's view of the goals of political community and understanding of fair distribution of material resources. However, what is clear is that in Athens – both during the reign of radical democracy in the 5th century and of more moderate democracy in the 4th – the well-off citizens were never in danger of having their land confiscated, their property taken away, or being taxed so heavily as to lose their wealth or status. As modern scholars have noted, the Athenian *demos* never used its collective power of decision-making to promote substantial redistribution of land and wealth (Ober 1996: 90–91). Contrary to Aristotle's claims in the *Politics* that the majority "will unjustly confiscate the property of the wealthy minority" (1318a25-26) or in the *Constitution of Athens* that "the democratic parties" are "rather in the habit of making a general distribution of land" (40), democratic Athens never resorted to redistributing land or confiscating property from members of the upper trierarchic class, not even after the defeat of Thirty Tyrants (Jones 1955: 153; Winters 2011: 87)[19]. Democratic government in Athens was never "dominated by the urban proletariat or the wider poor"; there was no radical redistribution of wealth, and democratic egalitarianism never translated into demands for social and economic equality, while prominent, affluent, and educated citizens were more likely to hold important public offices. The result was "a balanced modus vivendi between the social classes" (Mitchell 2015: 113).

[18] For this second point, see Demosthenes' speech *Against Meidias*.

[19] For a more detailed account of Athens' class structure, see: Winters 2011: 77–89.

It is worth noting that ancient claims about injustice in democracy refer to several different forms of injustice: the Old Oligarch detects it in the disregard for the natural superiority of the aristocratic class; Plato and Aristotle are concerned with democracy's neglect of true merit and excellence; and all of them are disturbed by the promotion of the interests of the poor majority at the expense of the wealthy minority, that is, by distributive unfairness. Once again, as revealed by both ancient sources and contemporary scholarship, the validity of these claims is provisional. They point to possible tendencies that exist in democracies, rather than establishing a persuasive argument about an innately unjust character of democratic rule. In conclusion, debates on the nature of democracy reveal that criticisms of popular rule were not completely unfounded: democracies can be inefficient, unstable, and unjust. However, they were not inherently so. One of the core features of democracy, which even many of its critics acknowledge, is its awareness of the dangers of demagoguery, *stasis*, popular tyranny, irresponsible decision-making, unjust redistribution, or imperial tendencies, and its ability to respond to these dangers by introducing institutional reforms and practices designed to curtail them.

References

Primary

Aristophanes. 1978. "Knights." In D. Barrett and A. H. Sommerstein (eds.). *The Birds and Other Plays*. London: Penguin Books, 29-90.

Aristotle. 1984. "Constitution of Athens." In J. Barnes (ed.). *The Complete Works of Aristotle, Volume Two*. Princeton: Princeton University Press, 2341–2383.

Aristotle. 1984. "Politics." In J. Barnes (ed.). *The Complete Works of Aristotle, Volume Two*. Princeton, NJ: Princeton University Press, 1986–2129.

Demosthenes. 1991. *Against Meidias*. Oxford: Clarendon Press.

Herodotus. 1996. *Histories*. Hertfordshire: Wordsworth Editions Limited.

Plato. 1997. "Apology." In J. M. Cooper (ed.). *Plato: Complete* Works. Indianapolis: Hackett Publishing Company, 17–36.

Plato. 1997. "Gorgias." In J. M. Cooper (ed.). *Plato: Complete* Works. Indianapolis: Hackett Publishing Company, 791–869.

Plato. 1997. "Letters." In J. M. Cooper (ed.). *Plato: Complete* Works. Indianapolis: Hackett Publishing Company, 1634–1676.

Plato. 1997. "Protagoras." In J. M. Cooper (ed.). *Plato: Complete* Works. Indianapolis: Hackett Publishing Company, 746–790.

Plato. 1997. "Republic." In J. M. Cooper (ed.). *Plato: Complete* Works. Indianapolis: Hackett Publishing Company, 971–1223.

Thucydides. 1972. *History of the Peloponnesian War*. London: Penguin Books.

Xenophon [Old Oligarch]. 1968. *The Constitution of the Athenians*. Cambridge: Harvard University Press.

Secondary

Balot, R. 2006. *Greek Political Thought*. Malden: Blackwell Publishing.

Cammack, D. 2013. "Aristotle on the Virtue of the Multitude." *Political Theory* 41 (2): 175–202.

Cartledge, P. 2007. "Democracy. Origin of: Contribution to a Debate." In K. A. Raaflaub, J. Ober and R. W. Wallace (eds.). *Origins of Democracy in Ancient Greece*. Berkeley: University of California Press, 155–169.

Farrar, C. 1988. *The Origins of Democratic Thinking: The Invention of Politics in Classical Athens*. Cambridge: Cambridge University Press.

Finley, M. 1985. *Democracy Ancient and Modern*. London: The Hogarth Press.

Hansen, M. H. 1991. *The Athenian Democracy in the Age of Demosthenes: Structure, Principles, and Ideology*. Oxford: Blackwell.

Hornblower, S. 1992. "Creation and development of democratic institutions in ancient Greece." In J. Dunn (ed.). *Democracy: The Unfinished Journey*. Oxford: Oxford University Press, 1–16.

Jones, A. H. M. 1953. "The Athenian Democracy and Its Critics." *The Cambridge Historical Journal* 11 (1): 1–26.

Jones, A. H. M. 1955. "The Social Structure of Athens in the Fourth Century." *Economic History Review* 8 (2): 141–155.

Loraux, N. 1986. *The Invention of Athens: The Funeral Oration in the Classical City*. Cambridge and London: Harvard University Press.

Maio, D. P. 1983. "Politeia and Adjudication in Fourth-Century B.C. Athens." *The American Journal of Jurisprudence* 28 (1): 16–45.

Mitchell, T. N. 2015. *Democracy's Beginning: The Athenian Story*. New Haven and London: Yale University Press.

Monoson, S. S. 1998. "Remembering Pericles: The Political and Theoretical Import of Plato's Menexenus." *Political Theory* 26 (4): 489–513.

Ober, J. 1989. *Mass and Elite in Democratic Athens: Rhetoric, Ideology, and the Power of People*. Princeton: Princeton University Press.

Ober, J. 1996. *The Athenian Revolution: Essays on Ancient Greek Democracy and Political Theory*. Princeton: Princeton University Press.

Ober, J. 1998. *Political Dissent in Democratic Athens*. Princeton: Princeton University Press.

Parker, V. 1998. "Tyrannos: The Semantics of a Political Concept from Archilochus to Aristotle." *Hermes* 126: 145–72.

Raaflaub, K. A. 1983. "Democracy, Oligarchy, and the Concept of the 'Free Citizen' in Late Fifth-Century Athens." *Political Theory* 11 (4): 517–544.

Raaflaub, K. A. 2006. "Thucydides on Democracy and Oligarchy" In A. Rengakos and A. Tsakmakis (eds.). *Brill's Companion to Thucydides*. Leiden and Boston: Brill, 189-224.

Robinson, E. W. 2011. *Democracy Beyond Athens*. Cambridge: Cambridge University Press.

Salmon, J. 1997. "Lopping off the Heads? Tyrants, Politics, and the Polis." In L. G. Mitchell and P. J. Rhodes (eds.). *The Development of the Polis in Archaic Greece*. London: Routledge, 60–73.

Waldron, J. 1995. "The Wisdom of the Multitude: Some Reflections on Book 3, Chapter 11 of Aristotle's Politics." *Political Theory* 23 (4): 563–584.

Winters, J. A. 2011. *Oligarchy*. Cambridge: Cambridge University Press.
Winton, R. 2005. "Herodotus, Thucyidides and the Sophists" In C. Rowe and M. Schofield (eds.). *Cambridge History of Greek and Roman Political Thought*. Cambridge: Cambridge University Press, 89–121.

Croatian Journal of Philosophy
Vol. XXVI, No. 77, 2026
https://doi.org/10.52685/cjp.26.77.4
Received: June 30, 2025
Accepted: January 21, 2026

In Defence of the Right to Out Others

MILOŠ KOVAČEVIĆ*
University of Belgrade, Belgrade, Serbia

Unlike visible, stigmatised personal characteristics, sexual orientation can be relatively successfully hidden. By staying in the closet, many queer people manage to minimise stigmatisation. However, intergroup contact is of key importance for increasing tolerance and liberalising the straight majority. The choice not to disclose one's sexual orientation thus leads to a collective action problem. Although the social stigmatisation of queer people could be more effectively overcome through mass coming out, suboptimal results are achieved owing to the strategic use of the right to privacy in passing as straight. This problem is most conspicuous in countries where the cost of coming out remains significant, while the cost of staying closeted is too low to trigger substantial, progressive change. I discuss the issue of the permissibility of outing through an analysis of both Jean Louise Cohen's reflexive conception of privacy and Judith Jarvis Thomson's reductionist conception of privacy. I argue against the understanding of the right to privacy as a right that prevents others from obtaining information about one's sexual orientation through legal means and from disseminating such information if it is not conditioned on confidentiality.

Keywords: Privacy; collective action; intergroup contact; sexuality; outing.

1. Introduction

It seems that philosophers rarely have the opportunity to straightforwardly contribute to the current public debate, as in the case of the outing controversies that have been happening in Croatia for several years. In the summer of 2022, the well-known Croatian journalist Mis-

* The realization of this research was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia as part of the financing of scientific research work at the University of Belgrade – Faculty of Philosophy (contract number 451-03-33/2026-03/200163).

lav Bago committed suicide. The day after queer activist Viktor Zahtila outed him, tweeting the following:

> I knew Mislav Bago superficially, but enough to know how deeply he suffered because he was in the closet. The last time we met, I was persuading him to come out. Not just to encourage young journalists/colleagues, but also because of personal happiness, self-confidence, and mental health. (@v_zahtila, August 19, 2022)

The reactions to Zahtila's tweet were overwhelmingly negative. The angry commentators, both queer and those who perceive themselves as allies, overwhelmed social networks: "You outed him by force!" "Why did you bring *that* up in public?!" "The closet was his *choice*!" etc.

Zahtila had a lot to unpack here. And he did. In his article, "Mislav Bago was being killed by Croatian homophobia. We cannot be silent about it," Zahtila noted that Bago's homosexuality was a public secret. He denounced attempts by the straight public to monopolise public memory of Bago and criticised the internalised homophobia of the queer community. Of special importance for my purposes is Zahtila's defence of public outing:

> Homophobes tell us to keep our sexuality "within our own four walls", *abusing terms such as "intimacy" and "privacy"* because violence is most easily reproduced in secret. *It is easiest to oppress someone when you know they will not rebel.* Therefore, there is no worse thing you can say to an LGBT person than that sexuality is their 'private matter' [...] we not only have the right but also the duty to do everything we can to encourage and help people to come out of the closet so that they can live in freedom and not rot in the dark [...] As an emancipated queer, I will not let yet another gay person be left in the darkness of death. A man should be remembered exactly as he was and not keep his homosexuality as a dirty secret that casts a shadow on his public image. (Zahtila 2022, emphasis added)

Two years later, two more high-profile outing cases followed. In the winter of 2024, President Zoran Milanović, apparently unwittingly, outed Minister Damir Habijan while responding to a journalist's question regarding a joke that had been made about Minister Habijan. Reacting to the chuckle that spread among the journalists after he disclosed the minister's sexual orientation, President Milanović seemed genuinely perplexed: "Habijan is gay, right? The man is not in the closet; he lives a normal life." President Milanović's reaction indicates that Minister Habijan inhabited a "glass closet", i.e., living authentically while refraining from any official declaration of his sexual orientation. Nevertheless, unwritten social rules create a strong expectation that everyone should participate in hiding the fact that he is gay, or at least remain silent about it. This sentiment was perhaps best expressed by one commentator criticising President Milanović, who stated that "no matter how much this [Minister Habijan's sexual orientation] was 'known' in certain circles, according to basic etiquette, such private information should not be publicly discussed" (Modrić 2024).

Not long after, the segment of the public critical of outing shifted from enforcing etiquette against transgressors to suspecting them of actually breaking the law. In other words, complaints about violations of unwritten rules were soon replaced by accusations of violating rules written in nothing less than the Criminal Code. Several months after President Milanović outed Minister Habijan, the major LGBTIQ civil society organisation Zagreb Pride outed several prospective members of parliament (hereafter MPs) during the 2024 election campaign by publishing an LGBTQ+ compass, i.e., a publicly available list of queer MP candidates. In an article titled "Zagreb Pride violated the legality of the election. This requires a response from the Constitutional Court and smells like a criminal offence", lawyer Vlaho Hrdalo argued that Zagreb Pride, by publishing the pre-electoral LGBTQ+ compass, violated Article 146 of the Criminal Code, which prohibits the illegal use of personal data of individuals (Hrdalo 2024).

I will discuss two questions. The first concerns conceptual confusion around the notion of *privacy* and its potential weaponisation against queer people. What does it mean to violate someone's privacy? Why would we bother sanctioning this type of violation? And on whom exactly should the burden of protecting privacy fall? The other question concerns the relationship between oppression and *secrecy*. Does secrecy make us more or less vulnerable? If the closet is so bad, why are so many queer people still not out?

My hope is that, through answering these two questions, I will at least offer some clarifications and a more nuanced picture of the increasingly politicised practice of outing, if not a persuasive enough justification for its legal permissibility. This is my argument:

1. Inequalities based on prejudices are unjust.
2. Intergroup contact reduces prejudices.
3. There is almost[1] no contact between heterosexual and queer people as such without the disclosure of sexual orientation.
4. If (1), (2), and (3) are true, anything that blocks the disclosure of sexual orientation perpetuates injustice against queer people.
5. The right to privacy, which does not allow outing as such, is not justified.
6. Therefore, the right to privacy, which does not allow outing as such, unjustifiably perpetuates injustice against queer people.

All premises seem non-controversial, save for premise (5). However, to explain the free rider problem faced by the queer community, I delve into premises (2) and (3) in the first part of the paper. In the second part of the paper, I focus on premise (5), building both on the influential Judith Jarvis Thomson's reductionist conception of privacy and on Jean Louise Cohen's more recent, reflexive conception of privacy that specifically targets intimate relations. Here, I want to explain what I mean when I say that the right to privacy, which does not allow outing,

[1] See footnote 4.

as such, is not justified. Privacy concerns can be paired with violations of other rights. For example, I can break into one's home and, through that independently illegal act, simultaneously violate one's privacy. I am not interested in these types of cases but only in *pure* cases of privacy infringement; that is, situations in which information about one's sexual orientation is not obtained through any illegal deed or used in a way that violates a previous agreement.

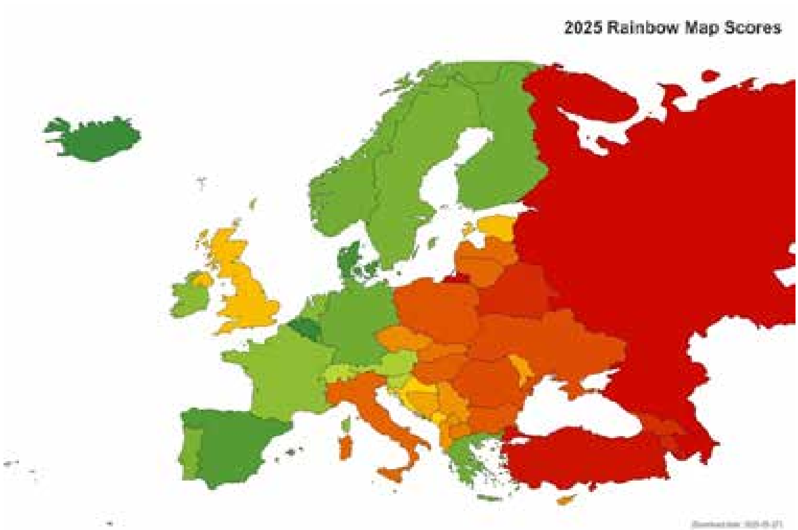
Let me briefly state what I am not defending to avoid potential objections. First, my argument is not about the moral permissibility of an outing. Importantly, there are acts that are morally wrong but nevertheless legal (e.g., lying). I am only defending freedom to out queer persons, but not a more demanding position according to which outing is morally justified or even a prudent thing to do in every single case and in each context. Therefore, legal permissibility in this context should be understood only as a juridical restriction on the right to privacy, but not as a legal duty to out queer people.

Second, I am not stating that intergroup contact is the only way to reduce prejudices. Although it seems to be one of the most effective, at least when we consider the achievements of the queer movement in the United States, I recognise that there are also other valuable ways of reducing prejudices, such as education or the employment of bias-correcting heuristics. One should also keep in mind that in highly prejudiced contexts, community-building is a prerequisite for any future activities that directly target prejudice against queer people. In the early stages of movement-building, queer communities can even benefit from privacy protections, as they encourage closeted individuals to participate in certain activities without fear of their sexual orientation being disclosed. However, as the movement matures, privacy protections increasingly become an impediment rather than an enabler of social change.

This dynamic leaves queer communities in "middle-equality countries" trapped in a way similar to how middle-income countries are stuck at their level of economic development[2]—unable to progress from anti-discrimination laws and same-sex unions to more robust legal protections and, more importantly, genuine acceptance by the heterosexual majority. Croatia, Serbia, and nearly all other former Yugoslav countries serve as examples of middle-equality countries in terms of queer freedom and equality.

[2] The middle-income trap refers to a situation in which a country experiences a significant slowdown in economic growth after achieving middle-income status, failing to transition to a high-income economy. This stagnation occurs because the country loses its competitive edge in labour-intensive, low-wage industries but struggles to compete in high-value, innovation-driven sectors. Essentially, rising wages make it difficult to compete with lower-income countries, while a lack of innovation and productivity hinders its ability to compete with higher-income countries.

One might imagine these countries as existing somewhere between Russia and Germany in terms of queer rights. They roughly correspond to the countries marked in yellow and orange on ILGA-Europe's Rainbow Map, which ranks European nations on a scale from red (gross human rights violations, discrimination) to dark green (respect for human rights, full equality) based on laws and policies directly affecting queer people's rights.



In these places, the cost of coming out remains too high, while the cost of staying in the closet is too low to trigger substantial progressive change, keeping the queer movement in stalemate. Although the closet is generally not a good place to be, some closets are better than others. While queer people in low-equality countries live in narrow, dark, and suffocating closets, those in middle-equality countries generally inhabit lighted, spacious, walk-in, and even glass closets. What does this mean in terms of real-life experience? It means the social pressure to pretend to be straight is much lighter, and thus the burden one must bear living in the closet is far less severe. There is no need for fake marriages or opposite-sex relationships. Individuals can live with a same-sex partner, and though it may raise some eyebrows if they continue to live with a 'roommate' after their twenties, few will confront them directly.

Perhaps most importantly, the process of disentangling homophobia from straightness is largely complete in middle-equality countries, so closeted queer people are protected from the most humiliating forms of self-betrayal. They do not have to perform casual homophobia to prove they are straight, as individuals in low-equality countries often must. In certain contexts, they can even be queer allies or have openly queer friends without anyone challenging their perceived 'straightness'.

This comfortable closet is facilitated by some legal protections, allowing individuals to behave more freely because they are less terrified of the consequences of being outed compared to those in low-equality countries. But even more crucial is the operative institution of the public secret. There seems to be a tacit understanding in these societies that queer individuals can live as they wish, provided they refrain from ever explicitly stating they are queer or publicly showing same-sex affection. Society, in turn, will wholeheartedly help them to remain in their comfortable closets, pretending not to see the obvious.

Third, many discussions of outing are limited to a special group of queer people. For example, in the case that I have described, part of the public discussion went in the direction of questioning whether dead people can have rights at all, including privacy rights. Other examples include the Barney Frank rule, according to which public outing is defended, but only for closeted right-wing politicians who use their power to harm the queer community. On the other hand, my argument in favour of outing is broader in scope and includes all queer people in middle-equality countries, regardless of whether they are dead or alive, celebrities or ordinary people, helping or harming the cause. However, I am excluding trans people from my discussion about the permissibility of outing because, for many of them, in contrast to lesbians, gays, and bisexual people, passing is important to being who they are (Overall and Sellberg 2016).[3] Therefore, I use the term "queer" as an umbrella term that refers only to all persons of a sexual orientation different from heterosexual.

2. Faustian bargain

In the book *The Nature of Prejudice,* Gordon Allport developed the hypothesis that contact between members of different groups under conditions of equal status, shared goals, and cooperation reduces prejudice among a normally prejudiced population (Allport 1954). Empirical findings support the idea that positive intergroup contact leads to a reduction in prejudice. In their meta-analysis, "A Meta-Analytic Test of Intergroup Contact Theory," Thomas Pettigrew and Linda Tropp examined hundreds of studies on intergroup contact and its effect on

[3] In principle, my argument in favour of the right to out others could also be applied to trans individuals who find the very notion of passing oppressive. This application, however, requires two specific conditions. First, the closeted trans person subjected to outing must hold that belief. Second, the country in question must qualify as a middle-equality country for trans people, not just for LGB people. In practice, this second condition is often not met, as the level of legal protection and general social acceptance for trans people is frequently significantly lower than for their LGB counterparts.

prejudice reduction. They found consistent evidence that positive contact between members of different groups was associated with reduced intergroup prejudice. Moreover, the same trends are observed in studies involving racial or ethnic groups as well as those involving different demographic groups. This finding implies that the contact hypothesis, initially developed for interactions between racial and ethnic groups, can also be applied to other groups (Pettigrew and Tropp 2006).

Employing the contact hypothesis in the context of homosexuality, Jeremiah Garretson, in his book *The Path to Gay Rights*, emphasised the importance of coming out and direct intergroup contact for the liberalisation of homophobic people (Garretson 2018). One of the most powerful tools for changing deep-seated prejudices is the act of coming out. When queer individuals disclose their sexual orientation, they are challenging preconceived ideas that prejudiced people have about them. Once the possibility of intergroup contact is unlocked, interactions between queer and straight people play a crucial role in reducing prejudice. Garretson emphasised that, owing to intergroup contact, in just 25 years, more than one-third of Americans completely changed their attitudes towards homosexuality. Positive attitudes towards queer people are critical for the advancement of their rights, as they create an environment where discrimination is less tolerated. Ultimately, the liberalisation of the straight majority is instrumental for legal change in support of equal rights for queer people, such as anti-discrimination laws and marriage equality. Thus, direct intergroup contact is a powerful way to address challenging prejudices. By strengthening personal connections, these processes pave the way for full equality between queer and straight people.

However, what blocks the liberalisation of the straight majority is not a lack of interaction with queer people, but rather the fact that their sexual orientation remains unknown. Even though intergroup contact undoubtedly exists, it cannot contribute to the liberalisation of the straight majority because they perceive it as interaction that occurs within their own group. Social stigma is normally the reason why queer people hide their sexual orientation. At the same time, staying in the closet maintains the power of prejudice because it prevents straight people from correctly interpreting their experiences with queer individuals and potentially changing their attitude for the better.

The vicious cycle of stigma that encourages staying in the closet, which in turn further maintains the stigma, persists due to the free rider problem. Prospective liberation for an oppressed group usually entails a certain cost for its members. We can look at it as an investment in a better future. By disrupting the *status quo*, we make a certain sacrifice now to live far better in the future.

Consider a simple model with three agents: A, B, and C. In an ideal collective action scenario, all begin at an equal baseline utility, contribute equal effort during a transition, and reap equal rewards from the resulting change.

In a free-riding scenario, the same collective good is produced but with a disproportionate distribution of costs. The entire transition cost can be borne by C alone, allowing A and B to retain their pre-transition utility without sacrifice. While this outcome is utility-maximising for A and B, it is clearly suboptimal for C.

Of course, social groups have many more members, and they also differ in the degree of utility they derive from the state of affairs that preceded the change. Additionally, it is possible that not everyone will benefit equally from the change. However, the inequality that interests me in this paper is the one that we find in the phase that divides oppression and liberation, i.e., in the phase of *change*. How is the burden of change shared?

In reality, change rarely depends on a single factor. An oppressed group may employ different liberatory tactics to achieve the same goal. For example, Garretson argues that indirect intergroup contact mediated by the media, more specifically, queer characters who were given a more prominent place in television shows, was another important factor in the liberalisation of the straight majority in the United States. Thus, it is possible to imagine a society with an extremely low level of coming out, but where the stigma towards queer people decreases due to the high consumption of queer-friendly media content.

The point is not that liberatory tactics unrelated to direct intergroup contact are ineffective at reducing prejudice towards queer people, but rather that they do not work equally well at every stage of queer liberation. This is especially visible in queer communities caught in the middle-equality trap. In the absence of widespread outing, old tactics yield diminishing returns until they can no longer produce significant change.

Within the context of the middle-equality trap, the contribution of other tactics to positive change approaches zero, so we can simplify our model. For the sake of analytical clarity, let us imagine that only direct intergroup contact drives the affective liberalisation of the straight majority.

As much as certain queer persons interact with straight individuals, their contribution to intergroup contact cannot be considered effective if heterosexual people are not aware of their sexual orientation; i.e., if they assume that they are also straight. Therefore, coming out is a *conditio sine qua non* of non-negligible intergroup contact and the consequent reduction of stigma.[4] In the absence of other sufficiently effective liberatory tactics, group members who do not meet this necessary condition are free riders who will reap the benefits of the change

[4] A baseline level of intergroup contact is inevitable, as certain queer individuals do not pass as straight. Nevertheless, depending exclusively on the interactions of this specific group would be ethically problematic and strategically ineffective for generating meaningful positive change.

without sacrificing for it. From an individual perspective, free riding is a rational behaviour because it avoids costs in the phase of change and achieves gains in the liberation phase. From a collective perspective, however, it is undesirable, but it can be tolerated if only a small part of the group resorts to it. Problems arise when free riding becomes the dominant strategy within the group, as the desired effects are not sufficiently produced, thus delaying liberation.

Free riding is not always as obvious as when a closeted queer individual has no political engagement that challenges homophobia. It can also take a highly active form, as when closeted individuals participate in a type of queer activism that no longer yields results. The reasons for this behaviour can be a genuine failure to recognise that their community is caught in a middle-equality trap, or a willful ignorance that rationalises remaining in the closet. This means that free riding does not have to be intentional. Many closeted individuals may mistakenly believe they are doing something meaningful to combat homophobia when, in reality, they are simply spinning their wheels.

The queer movement can encourage queer individuals to come out, explain its importance for the community as a whole, and highlight the personal benefits that it brings. However, such incentives are unlikely to be sufficient to overcome the grim reality of discrimination and violence that the individual is likely to face if they decide to take such a step. Queer people are stuck in making "a Faustian bargain: a duty of privacy (secrecy, concealment, silence, shame) in exchange for benign treatment (being let alone)" (Cohen 2002) and act in accordance with "the sentiments of many conservatives who see the closet merely as a way for homosexuals to avoid discrimination, a position consistent with the idea that anti-gay discrimination cannot be compared to other sorts of discrimination because it is avoidable through closeting" (McCarthy 1994: 32).

There is even more reason for pessimism if it is taken into account that the effectiveness of such incentives is inversely proportional to the intensity of homophobia in a particular society. This leads to a situation in which, in societies where coming out is most necessary, the queer movement has the weakest tools to encourage it. How do we break out of this deadlock?

Free riding can be countered by structuring the situation to preclude the possibility of non-cooperation altogether. States do that every day. For example, the state will not wait for everybody to recognise the importance of public goods to collect taxes. Can we apply the same logic to the discussed problem? Can we simply out free riders to speed up collective liberation? To answer this question, we need to examine whether outing ought to be legal.

Is outing a violation of a *right* or merely the failure to meet one's *preferences*? If the first is the case, then it is wrong because it infringes on an interest that is recognised as legitimate. If the latter is the case, then it is permissible because the nature of political life in a plural society is such that we have various conflicting preferences, and no group, be it marginalised or not, can expect a guarantee of all its preferences. In other words, preference satisfaction is the subject of legitimate political contestation.

The concept of privacy is often criticised for its lack of specificity (Solove 2006). While at the denotative level it serves to signify various interests, the connotative role of this term is persistent and serves to positively represent any of the numerous interests to which it allegedly refers (BeVier 1995). In a situation in which, in the words of Thomson, "the most striking thing about the right to privacy is that nobody seems to have any clear idea what it is" (Thomson 1975: 295), invoking privacy in connection with specific interests can serve to transfer a positive connotation from the concept of privacy to those interests, even though they have only contingent connections with it. In this way, mere personal preferences can be incorrectly inflated in importance, and failure to fulfil them can be misrepresented as a violation of the right to privacy.

I believe that the preference not to be outed is wrongly presented as a legitimate interest. The idea that preserving one's closet is somehow binding for the rest of us seems deeply problematic. All who have true information are effectively silenced through the imposition of the alleged obligation to hide the sexual orientation of closeted queer people. They are forced to perpetuate a false, biased, and harmful picture of the sexual orientation makeup of society, contributing to the further ossification of historic injustice (McCarthy 1994). That cannot be right. Let us see why.

3. Outing through the lens of a reflexive conception of privacy

In the book *Regulating Intimacy,* Cohen writes: "Privacy rights entail freedom from the obligation to conceal as well as informational privacy, i.e., the choice not to reveal or have revealed one's personal affairs. The correlative duty to such a right is that others respect it" (Cohen 2002). If the scope of the duty that stems from this definition of privacy is still unclear, Cohen explicitly adds in a footnote that it forbids both the enforced hiding of sexual orientation and *outing*. The equation that Cohen makes between the conservative imposition of secrecy on the queer community and the progressive rejection of privacy as an effective tool for fighting homophobia does not appear only in this particular place; rather, it looms throughout the entire chapter, in which she discusses the regulation of homosexuality. For Cohen, the imposition of secrecy and the rejection of privacy are nothing more than two sides of the same coin. Both positions are based on a misunderstanding of the right to privacy because they mistakenly take it as a duty of privacy, i.e., an obligation not to disclose one's queerness publicly. The only difference is that while conservatives embrace the duty of privacy as a condition for a "respectful" way of being queer, progressives reject it.

The obvious problem with the conservative conceptualisation of privacy is that it is partial, i.e., inconsistently applied to the majority and minorities. The double standards for queer and straight people are sometimes explicit, as in the "Don't Ask, Don't Tell" policy,[5] which targets only same-sex conduct, but they can also be covered up, presenting themselves as a general impartial standard. Cohen correctly noted that "the allegedly neutral civic self usually turns out to be filled with particular content determined and privileged by the majority" (Cohen 2002). What conservatives put on the table is an unfair deal, according to which, in return for the duty of secrecy, the queer community will obtain what Rainer Forst calls *the permission conception of tolerance*. According to this conception of tolerance, minorities are tolerated but only as politically and morally inferior to the majority. As long as they do not challenge the dominance of the majority, they can count on non-interference within the limited domain determined by the dominant group (Forst 2018). In contrast, Cohen is advocating for the full moral and political equality of minorities and *the respect conception of tolerance,* which is based on reciprocity, i.e., the willingness to offer reasons for regulation that both sides could reasonably accept. According to Cohen, the right to privacy accompanies the respect conception of tolerance and facilitates the transition from the permission conception of tolerance to the respect conception of tolerance.

As explained in the previous section, the critique of the progressive camp is that the focus on privacy depoliticises the oppression faced by the queer community and offers a false way out of repression (Mohr 1988). Cohen replies that this argument is not applicable to the right to privacy that she defends because it recognises the discretion of individuals over the choice of whether they will disclose their sexual orientation. According to her, if the right to privacy is correctly understood and coupled with a conception of tolerance on the basis of respect, there is no reason to be concerned that queer people will be silenced.

I agree with Cohen that, in the context of tolerance, the right to privacy can function beneficially for queer individuals, just as it functions very well for straight people now. However, the problem is that many queer individuals do not live in that kind of context. The trouble with the right to privacy is not that it *inherently* harms queer people in *any* circumstances, but that it is harmful *today,* at *this particular stage* of the struggle against homophobia.

If Cohen approached the issue of outing only from the perspective of an ideal theory, her defence would be acceptable. However, she clearly

[5] The 'Don't Ask, Don't Tell' policy was the official United States military policy from 1994 to 2011 regarding LGB service members. It was a compromise that barred military officials from investigating or asking about a service member's sexual orientation, while service members were prohibited from openly disclosing their homosexuality or engaging in homosexual conduct.

recommends her understanding of privacy for nonideal conditions as well. It is telling that Cohen defends the right to privacy in the phase of change, emphasising how even "if it is true that gaining the protective shield of privacy rights to cover gay sex will not of itself deliver full equality of power to homosexuals, the denial of this shield has played and continues to play a central role in supporting official and unofficial oppression, surveillance, and discrimination against them" (Cohen 2002). Similarly, she adds, even if the right to privacy "(...) will not yield the full panoply of rights that gays and lesbians need (...) it would pull the rug out from under the rhetorical legal construction of gays and lesbians as a tendentially criminal population undeserving of the rights that others enjoy" (Cohen 2002). Thus, according to Cohen, the right to privacy is preferable both at the end-state of full equality, because it enhances personal autonomy, and at the transitional state, because it lowers the costs of resisting oppression.

It seems that Cohen is conflating lowering the costs of *resisting oppression* with lowering the costs of *giving up*. If there are tools that can lower the costs of confronting injustice while still effectively resisting it, they are always welcome. However, that is a totally different thing from simply quitting the confrontation through a cover-up and then, in addition, imposing an unconditional legal obligation on others to tacitly support one in the withdrawal. Instead of being a tool for empowering changemakers, this rather looks like a recipe for their failure. I have explained in the previous section how the right to privacy causes very concrete harm to the queer movement by enabling free riding. It is not enough to show that, in a particular case, some queer individuals can use the privacy right to shield themselves against homophobia to argue that privacy helps queer liberation. It should be shown that:

1. The queer individual does not protect herself at the expense of the group's interest;
2. The benefits that the right to privacy brings outweigh all costs that it imposes on the queer community as a group, especially its facilitation of mass free riding.

In the context of middle-equality countries, queer individuals rarely use their right to come out and predominantly decide to stay in the closet. Therefore, even if they do not have a *de jure* duty of secrecy, they have one *de facto*. Queer people end up with equality only on paper if their real-life circumstances are not taken into account. In one way, Cohen presupposes toleration on the basis of respect that has yet to be established. The other problem is that she states that the right to privacy will be helpful in queer liberation, offering a justification that does not separate personal benefits from collective benefits and that does not make an overall cost–benefit analysis of the criminalisation of outing.

Am I begging the question here? If there is a right against being outed, then it seems that the cost–benefit analysis is misdirected since the rights cannot be violated, even if that would bring a greater overall benefit. This is generally true once the right is established, but here I am discussing whether such a right should exist at all. For the introduction (or contestation) of any right, a cost–benefit analysis is crucial.

One could respond that even if the closet is what queer people choose, that is all right, because the whole point of the right to privacy is to protect the personal autonomy of individuals. If the closet was Bago's *choice*, as one commentator wrote, his autonomy should be respected. There are choices that the queer movement would probably not like, but that are nevertheless autonomous and should be protected by the right to privacy.

To answer this objection, we need to dip into the muddy waters of personal autonomy and raise a difficult question about how much we can rely on the expressed preferences of oppressed people, especially given their adaptability to unfair and distorting social conditions. Broadly speaking, we differentiate between internalist and externalist accounts of personal autonomy (Colburn 2022). While internalist accounts emphasise the importance of alignment between lower-order and higher-order preferences, sidelining the conditions in which identification occurs, Marina Oshana's externalist socio-relational account of personal autonomy denies the authenticity of the expressed preferences of oppressed agents. According to Oshana, if individuals are not independent from others in a way that enables them to effectively contest and block arbitrary influences without placing an unreasonable burden on themselves, then their choices cannot be considered autonomous (Oshana 2016). It is clear that the opportunity to easily and effectively challenge arbitrary influences does not exist for many queer people, at least when it comes to issues regarding their sexual orientation. Moreover, the level of independence inversely correlates with the actual need for an outing. In societies where homophobia and transphobia are particularly strong, we can expect that the autonomy of queer people is particularly compromised; therefore, the authenticity of their expressed preferences is even more unreliable.

One might question the need for such strong safeguards against arbitrary influences, as closeted individuals, by not disclosing their sexual orientation, avoid direct exposure to homophobia. This could foster the mistaken belief that their choice renders them less vulnerable and more autonomous, even in hostile settings.

However, the republican nature of the independence that is required for personal autonomy does not allow oppression to be wished away. In contrast to a liberal understanding of freedom, republicans demand that free people have certain options open to them, even if they do not actually use them. Therefore, if one cannot block arbitrary influence without undue burden, they are not independent, even if they are in the closet and not even planning to come out.

One could still argue that even if queer people are not autonomously deciding to stay in the closet, that does not make outing them permissible. Consider an armed robber who orders me to give him my money under threat of being shot. Since this is a textbook violation of autonomy, one cannot plausibly argue that my decision to hand over my wallet is autonomous. Nevertheless, this does not mean a bystander should stop me from doing so and risk my life in that way. In the same way, a third party should not interfere with an individual's decision to remain closeted, because that decision, however non-autonomous, preserves their livelihood.

There is some merit to this counterargument. On an individual level, a closeted life is obviously better than death, and it is difficult to imagine a collective benefit high enough to compensate for the loss of human life. Yet this is precisely why a cost–benefit analysis is so crucial for properly addressing the issue, and why I limit the scope of my argument to middle-equality countries. In my view, the right to out others is redundant in high-equality countries and too dangerous in low-equality ones. The thought experiment with the armed robber reflects the typical experience of queer individuals in Chechnya, but not in Croatia. What is also missing from the example is the presence of other people whom the robber is threatening, and a description of how my cooperation negatively affects their situation.

A better model for a middle-equality country would be a schoolyard bully who demands lunch money from everyone, threatening social torment rather than a shooting. Some victims endure the bullying, others comply to avoid it. Here, the act of handing over money directly resources the bully, and, more critically, each act of compliance signals to the wider group that his power is legitimate. This perception makes his harassment of those who resist more frequent and severe. In this scenario, it is far less clear that a bystander should refrain from intervening. But even if we conclude that a bystander lacks the right to interfere, it does not follow that other victims must also refrain. On the contrary, those actively being tormented for their resistance have the right to stop those whose compliance enables their torment.

4. Outing through the lens of a reductionist conception of privacy

Thomson argued that every attempt to find a necessary and sufficient condition for a violation of the right to privacy fails (Thomson 1975). According to Thomson, the right to privacy is best understood as a derivative right; that is, a type of right that always overlaps with other rights—e.g., the rights to property, life, and liberty, and the cluster of "un-grand rights" that encompass minor things that can be done with our bodies, such as cutting one's hair while they sleep, touching one's knee without consent, and similar. These are not normally considered violations of the right to bodily integrity, which Thomson calls "the right over the person" (Thomson 1975: 305). Although it is possible to violate both grand rights and un-grand rights without violating the right to privacy, it is impossible to violate only the right to privacy, i.e., without violating some other right.

To argue for a derivative character of the right to privacy, Thomson constructs several thought experiments through which she examines various possible scenarios of its violation. One that seems particularly useful for our discussion is one about Professor Jones, the owner of a pornographic picture. Thomson recognises the right to property, which allows Professor Jones to control who can look at the picture and, in that way, protect his privacy. Therefore, it would be a violation of Professor Jones's right to privacy if one gained knowledge about the picture by stealing it, spying on him, threatening him, or torturing him. However, if Professor Jones waived his right to stop someone from watching his pornographic picture, his privacy would not be violated. For that reason, Thomson suggests a shift of focus from defining the right to privacy to detecting other violations of rights. If we adequately protect other rights, the right to privacy will also be automatically protected because its violation is always a result of infringement of other rights.

The right to privacy can be violated both at the input and output levels. More precisely, the right to privacy can be violated on both ends of the process of *knowing something*, but the possession of information itself is not a violation of any right (Persson and Savulescu 2012). On the input level, i.e., the level of acquiring information, if I did not threaten you, torture you, spy on you, or violate any other of your rights to find out that you are queer, I have not violated your right to privacy either. Let us say that you freely told me about your sexual orientation or that I saw you kissing a same-sex person at a party. I have not violated your right to privacy because hearing what people tell me and looking at people at parties do not violate any right.

However, when we speak about the legality of outing, concerns are normally not about how someone acquired information about another person's sexual orientation. After all, no one accused Zahtila of spying on Bago. Nor did Hrdalo imply that Zagreb Pride committed a criminal offence by illegally acquiring the personal data of closeted prospective MPs. They were concerned with the output level, i.e., how information was *used*.

Just as there are cases in which the way of *acquiring* information violates one's privacy, it is also possible to *use* the information in ways that would also violate the right to privacy. At the output level, breaching confidentiality is a way to violate the right to privacy. If one freely tells me that they are queer, they can, in advance, condition the dissemination of that information on confidentiality. Otherwise, I have no default obligation to "cooperate with the closet," even if outing could lead to others harming the outed person (Mohr 1992; McCarthy 1994: 37). In the cases where I simply saw someone kissing a same-sex person at the party, there is obviously no obligation to hide that information either. Thomson makes it clear that if an obligation of confidentiality does not exist, one should be allowed to speak freely and publish information acquired in a legitimate way (Thomson 1975).

Rights can be waived, both intentionally and unintentionally. The latter type of case is of special interest to us because it cannot be conditioned on confidentiality. Unintentional ways of not claiming the right include cases where, although the person who held the right did not intend to waive it, they nevertheless did so accidentally, due to negligence.

If we make the reasonable assumption that all cases of intentionally waiving the right to privacy in regard to information about sexual orientation will be made conditional on confidentiality by closeted people, then it seems that the issue of the legality of the outing boils down to the question of what counts as unintentional waiving of the right to privacy. Thompson says that there are many factors that should be considered when estimating whether a right has been waived, including the following:

1. How important is the right?
2. How costly is it to secure the right?
3. What are the customs and conventions?

In the case of (1), the important difference is who is actually the holder of the right. If the holder of the rights is an autonomous person, then the right to privacy is important to her because it facilitates her personal autonomy. However, if one's autonomy is compromised, this right has little value. Privacy will then be invoked only to perpetuate and ossify preferences produced by illegitimate influences. As I have already shown, there are good reasons to believe that, owing to homophobic oppression, queer people do not enjoy adequate independence for their choices about coming out to be considered autonomous. Therefore, the right to privacy in this domain is not particularly important for queer people because it does not necessarily protect interests that can be justifiably recognised as their own. Nevertheless, the importance of the right to privacy could still be defended on paternalistic grounds, emphasising how it can shield queer people from harm. However, it is important to separate outers from the people who harm queer people. The freedom of speech of the outer and the right to a dignified life of queer people who are outed without permission are not necessarily in conflict. They are only in conflict if we wrongly suppose that the outer is responsible for the homophobic behaviour of other people (Mohr 1992). However, that would be a mistaken imputation of responsibility that violates the basic rule that every capable adult is responsible for their own behaviour. An authentic concern for the well-being of queer people requires protecting them directly from violence and discrimination through the effective prosecution of perpetrators, not indirectly through infringement of freedom of speech.

With respect to factor (2), Thomson is clear about the need to find the middle ground between excessively stringent and insufficiently rigorous security measures. On the one hand, there are privacy-protecting precautionary measures that put an unreasonable burden on the person. On the other hand, there are also cases in which one obviously has not done enough to protect their privacy. For the first type of case, she says that it would be unreasonable to expect that Professor Jones pays for a platinum wall safe just because an imaginary X-ray device (which, according to Thomson's thought experiment, could be used to look at his picture) cannot penetrate this very expensive material.[6] Regarding the second type of case, in which the right-holder has not done enough to protect their privacy and, in that way, has waived their right, Thomson illustrates it with the example of a couple having an argument next to the open window of their flat, so that passers-by cannot help but hear what they are saying.

Now, the question is whether the situation in which I see someone kissing the same-sex person at a party[7] is more similar to the case of Professor Jones, who cannot afford a platinum wall safe and in that way enables the owner of the X-ray device to see his pornographic picture, or to a couple who are having an argument next to an open window and in that way enables passers-by to hear what they are saying.

Before I delve into this question, I want to point out a tension between Thomson's factor (2) and the problem that I am addressing. The whole point of my critique of privacy is that it makes staying in the closet easier. In other words, extensive privacy protection creates a low-cost closet and, in that way, slows queer liberation. However, if the high costs of precautionary measures are a contributing factor that limits our understanding of waiving the right, then only utterly careless acts will be plausible indicators that information about one's sexual orientation may be freely disseminated by third parties. For that reason, it is difficult to see how factor (2) would make staying in the closet a less attractive choice. If anything, it makes it more appealing because it limits the effort one needs to put into protecting their privacy. Although the low-cost closet is compatible with Thomson's understanding of waiving the right to privacy, we should bear in mind the derivative nature of the right to privacy and that Thomson's factors—which are supposed to govern our estimation of waiving—do not apply in cases where no other right is violated.

This raises the question of whether looking at someone at the party is more similar to the couple fighting next to the window than to Professor Jones, who cannot afford a platinum wall safe. Thomson, rather controversially (Scanlon 1975), believes that people have the right not to be looked at or listened to. They are part of the un-grand rights that she recognises. Although I agree with Thomson that we should have these rights, I believe that the distribution of the burden of protecting any particular un-grand right will depend on the very nature of that right. This means that the distribution of the burden of protecting the

[6] More radically, Savulescu and Persson even argue that there is no state obligation to protect the privacy of citizens, even in cases in which the personal costs of securing privacy are very high.

[7] It is not known how Zagreb Pride activists discovered that the closeted MP candidates are queer, but we can imagine that they did so in this way.

right not to be looked at or listened to is quite different from that in the case of the right not to be touched. I believe that this is the case for two reasons.

The first is the libertarian idea of self-ownership, i.e., that we possess our bodies in the same way as we possess other things. Thus, just as you can watch my house, but you cannot enter without my permission, you can also watch me, but you cannot touch me.

The other reason is the different levels of intrusiveness of touching compared with listening and looking. Although they can all cause belief-induced distress, only touch entails an objective foundation—i.e., physical contact. Touching, unlike listening and looking, involves a factual, non-subjective change in the state of the person who experiences it. The objectivity of the change in the state of a person who is touched demonstrates that it is truly difficult to touch people without their noticing, whereas you can easily look at or listen to them without their awareness.

Indeed, we treat touching differently from looking and listening in everyday life. Consider how the burden of privacy protection is asymmetrically split. I can simply demand that you not touch me without any particular explanation, and there is indeed a strong social expectation—which could lead to a possible accusation of harassment—that you withdraw from such an act without question. On the other hand, asking someone not to listen to you or not to look at you would not only seem strange and rude but also unfairly infringe on their freedom. This is the reason why, in the case of listening and looking, the burden is distributed differently than in the case of touching. If I do not want to be listened to, I need to whisper. If I do not want to be looked at, I need to turn off the camera during a video meeting. The costs of securing privacy are put on the right-holder in regard to listening and looking.

I am not saying that there is no right not to be looked at or listened to. If someone were to demand that I turn on the camera on a video call, I would consider that an infringement of my privacy. However, that is a totally different thing from expecting others not to look at us when we have not even tried to hide ourselves. That is exactly what is happening when we treat seeing a same-sex couple kissing at a party as a violation of their privacy. Thus, although factor (2) is inherently antagonistic to queer emancipation, the most common cases of finding out about someone's sexual orientation do not involve placing an extraordinary burden on the right-holder to secure his privacy.

Factor (3), customs and conventions, is so vague and inherently conservative that it is of little use in this—or any other—discussion about privacy. Even if it is accepted as a factor that should govern our understanding of waiving rights, we must pay close attention to the level of generality at which it is interpreted. One cannot look at customs and conventions about homosexuality *per se,* as this would risk including homophobic norms in discussions. Nor is it acceptable to treat customs and conventions about sexuality in general as relevant, because this sphere of life is treated unjustifiably differently from others owing to sex-negativity. It is only plausible to take customs and conventions about privacy in the broadest sense, but that is precisely what is at stake in the entire discussion. The conventions are made and remade in various ways, including through rational public deliberation. Granting existing customs and conventions around privacy legitimacy, regardless of their content and the way they are generated, maintains a *status quo* that is not necessarily just. Thus, factor (3) blocks social change through which customs and conventions could be re-formed in a more reasonable way.

5. Conclusion

The desirability of a particular choice is always context-dependent, so it makes little sense to speak of staying in the closet as a categorically desirable or undesirable choice at the individual level. On a collective level, staying in the closet is not conducive to queer liberation. The reflexive conception of privacy puts hope in supplementary anti-discriminatory legislation that ought to transform the current duty of secrecy into merely one option among others—a choice queer people could make without penalty. However, it fails to provide an explanation for how robust anti-discriminatory protections could be achieved without mass outing. Coming out is not only a valuable end goal but also one of the most effective ways to reach it. The queer movement could rely only on other means to achieve full equality, but it is unclear why it should make such a concession. This is especially true in middle-equality countries, which require widespread direct intergroup contact to attain full equality and where the personal cost of being out is not, on average, unduly high. Since queer individuals cannot make the independent choice of whether to come out in homophobic contexts, the choice we face is not between personal autonomy and heteronomy but between two types of heteronomy—one that maintains injustice and the other that strives for equality. If we accept that inequalities based on prejudices are unjust—including those based on sexual orientation—then it follows that we should choose liberatory heteronomy over that which perpetuates the *status quo*.

When the right to privacy is interpreted as a derivative of other rights, the issue reduces to finding the correct measure for allocating the burden of securing the right, both in terms of who bears the task and how difficult it should be. The analysis of un-grand rights as well as the conditions for waiving the right shows that the alleged obligation of others to conceal the sexual orientation of queer people places an unreasonable burden of securing the right on everyone except the right-holder. The right to private property, the right over one's person, and other rights protect individuals from the most intrusive ways of acquiring information about their sexual orientation. Similarly, confidentiality limits the dissemination of acquired information. However, these rights cannot be overextended to include cases in which the right-holder has taken no precautionary measures to avoid being seen or heard. Therefore, the right to privacy is weaponised against the queer community when it is interpreted to block the outing of queer people, except in cases in which information about their sexual orientation is acquired through a rights violation or is used in ways that breach an agreed-upon confidentiality.

References

Allport, G. W. 1954. *The Nature of Prejudice*. Cambridge: Addison-Wesley.

BeVier, L. R. 1995. "Information about Individuals in the Hands of Government: Some Reflections on Mechanisms for Privacy Protection." *William & Mary Bill of Rights Journal* 4: 455–506.

Cohen, J. L. 2002. *Regulating Intimacy*. Princeton: Princeton University Press.

Colburn, B. 2022. "Introduction." In B. Colburn (ed.). *Routledge Handbook of Autonomy*. New York: Routledge, 1–10. https://doi.org/10.4324/9780429290411-1.

Forst, R. 2018. "Toleration, Justice, Reason." In C. McKinnon and D. Castiglione (eds.). *The Culture of Toleration in Diverse Societies*. Manchester: Manchester University Press, 201–222.

Garretson, J. J. 2018. *The Path to Gay Rights*. New York: New York University Press.

Hrdalo, V. 2024. "Zagreb Pride je prekršio zakonitost izbora. To zahtijeva reakciju Ustavnog suda i miriše na kazneno djelo." *Telegram*. https://www.telegram.hr/komentari/zagreb-pride-je-prekrsio-zakonitost-izbora-to-zahtijeva-reakciju-ustavnog-suda-i-mirise-na-kazneno-djelo/ Accessed 6 December 2025.

McCarthy, J. 1994. "The Closet and Ethics of Outing." In T. F. Murphy (ed.). *Gay Ethics: Controversies in Outing, Civil Rights, and Sexual Science.* New York: The Haworth Press, 27–45.

Modrić, S. 2024. "S Milanovićevim nasilnim iznošenjem Habijanove privatnosti nema veze ni HDZ, ni rivalstvo s Plenkovićem. To je naprosto sirova prljavština." *Telegram*. https://www.telegram.hr/komentari/s-milanovicevim-nasilnim-razotkrivanjem-habijanove-privatnosti-nema-veze-ni-hdz-ni-rivalstvo-s-plenkovicem-to-je-naprosto-sirova-prljavstina/ Accessed 6 December 2025.

Mohr, R. D. 1988. *Gays/Justice: A Study in Society, Ethics and Law*. New York: Columbia University Press.

Mohr, R. D. 1992. *Gay Ideas: Outing and Other Controversies*. Boston: Beacon Press.

Oshana, M. 2016. *Personal Autonomy in Society*. New York: Taylor & Francis.

Overall, C. and K. Sellberg. 2016. "Transgender Identity and Passing Authentically." In K. Harrison and D. R. Cooley (eds.). *Passing/Out*: *Sexual Identity Veiled and Revealed*. New York: Routledge, 203–232.

Persson, I. and J. Savulescu. 2012. *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.

Pettigrew, T. F. and L. R. Tropp. 2006. "A Meta-Analytic Test of Intergroup Contact Theory." *Journal of Personality and Social Psychology* 90 (5): 751–783. https://doi.org/10.1037/0022-3514.90.5.751.

Scanlon, T. 1975. "Thomson on Privacy." *Philosophy and Public Affairs* 4 (4): 315–322.

Solove, D. J. 2006. "A Taxonomy of Privacy." *University of Pennsylvania Law Review* 154 (3): 477–564. https://doi.org/10.2307/40041279.

Thomson, J. 1975. "The Right to Privacy." *Philosophy and Public Affairs* 4 (4): 295–314.

Zahtila, V. 2022. "Mislava Bagu je ubijala hrvatska homofobija. Ne smijemo šutjeti o tome." Index.hr. https://www.index.hr/vijesti/clanak/mislava-bagu-je-ubijala-hrvatska-homofobija-ne-smijemo-sutjeti-o-tome/2400688.aspx. Accessed 26 January 2024.

Croatian Journal of Philosophy
Vol. XXVI, No. 77, 2026
https://doi.org/10.52685/cjp.26.77.5
Received: February 11, 2026
Accepted: June 30, 2026

The Transgender Experience and the Limits of Physiological Continuity Theories

SAMUEL KAHN
Indiana University, Indianapolis, USA

In this paper I want to address an issue at the intersection of the philosophy of personal identity and the philosophy of transgender identity: the fact that, on one influential way of framing the transgender experience, it is incompatible with, and therefore constitutes a challenge for, physiological continuity theories of personal identity. To see this, I begin by outlining the relevant part of physiological continuity theories. Then I turn to the relevant thesis about the transgender experience. After explaining why these are in tension, I conclude with two alternate ways of moving forward, arguing that, although both approaches face challenges, either suffices to avoid this tension.

Keywords: Transgender identity; personal identity; physiological continuity.

In this paper I want to address an issue at the intersection of the philosophy of personal identity and the philosophy of transgender identity: the fact that, on one influential way of framing the transgender experience, it is incompatible with, and therefore constitutes a challenge for, physiological continuity theories of personal identity.[1]

Before I begin, I should make clear that the challenge is not raised uniquely by the transgender experience. But the transgender experience makes the challenge easier to see and, for that reason, harder to deny. I also should make clear that others might think that the challenge is really for transgender identity, or at least for the particular way of framing it that will be explored below, rather than for physiological continuity theories of personal identity. It is a matter of which

[1] Throughout, claims of incompatibility should be understood as restricted to the metaphysically robust reading specified below.

intuitions are more firmly held, and the point is that there is a core thesis in physiological continuity theories that is in tension with a core thesis about the nature of the transgender experience.

To see this, I begin by outlining the relevant part of physiological continuity theories. Then I turn to the relevant thesis about the transgender experience. After explaining why these are in tension, I conclude with two alternate ways of moving forward, arguing that, although both approaches face challenges, either suffices to avoid this tension.

Physiological continuity theories say that one's personal identity is tied to one's physiology in some important way (Kahn 2014: chapter 5; Olson 2024: section 7). Of course, physiology changes through time: red blood cells are being broken down in the liver and generated in the bone marrow; white blood cells are in a constant state of flux; epithelial cells are continuously being replaced; neural networks are formed, stabilized, destabilized, and stabilized again; and so on. But proponents of physiological continuity theories argue that (i) there is some rough continuity between the states of one's body from moment to moment, and (ii) this continuity is explained by some regular causal process. This is what is taken to explain personal identity through time.

Different versions of physiological continuity theory differ in important ways. For example, Sauchelli argues that we are essentially material objects and that these material objects can go through different phases, including those of human animal, person, and corpse (Sauchelli 2017). By way of contrast, according to Wiggins, once we die, we cease to exist: a corpse is what remains of us on Wiggins' account, not a phase of us (as on Sauchelli's account) (Wiggins 2003).

However, what these different versions of physiological continuity theory have in common, and what is most important for present purposes, is the assertion that we are identical to our bodies in some important way and, therefore, that it is not merely impossible, but incoherent, to suppose that we and our bodies could have importantly different properties of the sort that will be at issue in this essay (i.e., sexed properties central to identity, where "sex," as used here, refers to biological or physiological classification).

Physiological continuity theories are contrasted most frequently with psychological continuity theories. The latter approach personal identity in a similar way, as being tied to a process rather than a substance (unlike ensoulment theories of personal identity). But the process that identity tracks is, according to psychological continuity theories, a mental one rather than a bodily one: whereas physiological continuity theories tie identity to physiology, psychological continuity theories tie identity to psychology—to minds (Kahn 2014: chapter 5; Olson 2024: section 4). There is some overlap here insofar as physicalist theories of mental states will say that mental states are instantiated in the body. But the overlap is incidental and, more, physiological continuity theories focus on the body as a whole rather than solely those parts responsible for mental activity (primarily the brain).

To understand the differences between physiological and psychological continuity theories, it is helpful to consider some of their key points of contrast.

One argument made against psychological continuity theories appeals to the Birth Problem:

> If psychological continuity theories are correct, then we were never born. (Kahn 2014: chapter 5; Olson 2024: section 7)

The idea here is that the kind of robust psychological properties required for psychological continuity are not realized, according to empirical psychologists, until age 1 or 2, long after both conception and birth. Thus, if psychological continuity theories are correct, then we emerged into being about a year after our bodies were born and about two years after our bodies were conceived.[2] Because this is counterintuitive, it is a problem for psychological continuity theories—one which physiological continuity theories do not face.

A second argument made against psychological continuity theories appeals to what David DeGrazia has dubbed the Someone Else Problem:

> If psychological continuity theories of personal identity are correct, then deciding what medical decisions should be made in the event that one becomes psychologically incapacitated is deciding about what should happen to someone else. (DeGrazia 1999)

The Someone Else Problem is taken to show that psychological continuity theories undermine the authority presupposed in the practice of living wills, and, because the practice of living wills is thought to be unproblematic, the Someone Else Problem, like the Birth Problem, is supposed to pose a challenge to psychological continuity theories, but not to physiological continuity theories.

On the basis of these kinds of arguments against psychological continuity theories, some proponents of physiological continuity have begun to voice disapproval for brain death as a criterion of death (Pellegrino et al 2008). In cases in which integrative bodily functioning is still present but brain activity is irretrievably absent, legal codes (like those in the US and Europe) which honor brain death (of some kind) as a sufficient condition of death pronounce a person dead. But, according to these critics, because, in such cases, physiological continuity continues even when psychological continuity does not, pronouncements of death are premature.

[2] Although EEG activity can be detectable as early as 6 weeks, clear state transitions do not seem to appear until around 30 weeks (Britton et al 2016). This suggests that, even if there is an answer to the Birth Problem (if psychological continuity can be built on less robust psychological properties), it would founder on the idea that we were conceived, or that the first quickening event was when my mother first felt me move.

The Birth Problem, the Someone Else Problem, and questions about criteria of death concern either the beginning or the end of life of an individual. But the transgender experience is not generally framed in those terms. The transgender experience is generally framed as arising in the course of an individual's life. On one popular narrative—which will be challenged below—a member of the transgender community is someone who is sufficiently psychologically mature to have a concept of self; to be aware of their body; to be aware, in particular, of their body as sexed; and to feel a profound sense of alienation from their body insofar as it is sexed. That is, according to many individuals in the transgender community, transgender issues arise when the subject feels that they do not fit in their body (Erickson-Schroth 2022). Call this the Popular Transgender Narrative:

> Members of the transgender community believe that the sex of their bodies does not match their experienced sex.[3]

For the purposes of this paper, I am focusing on readings of such claims that treat "mismatch" as involving a difference in what the subject is (in the identity-constituting sense), rather than merely how the subject is socially classified or experienced. I should emphasize that this is a philosophically loaded reconstruction of the narrative: many individuals and theorists understand such claims in less metaphysically committal ways, for example as concerning gender identity, social role, or lived experience, rather than identity in the strict numerical sense, and I am going to explore precisely such an alternative reconstruction at the end of this essay. With that being said, the reason the Popular Transgender Narrative is important for present purposes is that it is predicated on a distinction between body and individual, a distinction that leads to two categories of individuals: cisgender individuals, whose bodily sex matches the sex with which they self-identify (i.e.,

[3] Moschella provides evidence that this is a popular transgender narrative (Moschella 2021: section II; see also Skrzypek 2023, esp. section IV). She uses this to point us toward the same tension between the transgender experience and physiological continuity theories that I am going to point us toward. However, Moschella goes further than I: she argues, in addition, that non-psychological continuity theories are mistaken; that transgender individuals are laboring under an illusion; and that "offering hormonal and surgical gender reassignment treatments for gender dysphoria is analogous to offering liposuction for anorexia" (Moschella 2021: 795). This analogy is controversial and not widely accepted, but it illustrates the strength of the position she defends.

As will emerge below, my goals are much more modest than Moschella's: I merely want to point to the tension between the Popular Transgender Narrative and physiological continuity theories; I am not going to take a stand on the probity of other theories of personal identity, and, as I will point out at the end of this essay, I remain open to ways of conceptualizing the transgender experience which are consistent with physiological continuity theories. Thus, unlike Moschella, I remain open to (a) retaining the Popular Transgender Narrative and embracing alternate theories of personal identity, and to (b) embracing physiological continuity theories but reconciling the way we think about the transgender experience therewith.

with their self-conception or psychological identification), and transgender individuals, whose bodily sex does not match the sex with which they self-identify.

I am not claiming that the phenomenology described above entails a metaphysical conclusion on its own; rather, it is relevant insofar as it is compatible with, and often naturally expressed in terms that suggest, such a reading. In support of this strong interpretation, it is worth attending closely to the way in which some transgender individuals describe their experiences in first-person terms. In a range of autobiographical and quasi-clinical contexts, individuals report, not merely that they are treated incorrectly, or assigned to the wrong social category, but that their bodies themselves feel wrong for them in a more direct and immediate sense. For example, descriptions of distress focused on particular sexed features, such as primary or secondary sex characteristics, are often framed not as complaints about how those features are socially interpreted, but as expressions of alienation from those features as parts of one's own body (Serano 2007; Mock 2014; Erickson-Schroth 2022). First-person narratives likewise frequently describe a sense of inhabiting a body that does not fit one's sense of self, rather than merely occupying an ill-fitting social role (Bornstein 2016; McBee 2014). While such statements are certainly open to multiple interpretations, one natural way of understanding them is as expressing a form of self/body mismatch that is not exhausted by claims about social role or presentation. At the very least, these cases make it difficult to rule out, from the outset, a reading on which the speaker is expressing a tension between what they take themselves to be and the sexed character of their body.

This point can be reinforced by considering accounts of gender dysphoria that emphasize the phenomenology of embodied distress. Clinical and theoretical discussions note that such distress is often directed at the body as such, rather than solely at the social meanings attached to it (American Psychiatric Association 2022; World Professional Association for Transgender Health 2022). Individuals frequently report discomfort, anxiety, or even revulsion in response to their own sexed anatomy, independently of any immediate social interaction or role-based constraint (Erickson-Schroth 2022). Phenomenological and philosophical treatments likewise stress the immediacy with which the body can be experienced as alien or discordant with one's sense of self (Heinämaa 2003). If this phenomenology is taken at face value, then a purely social or role-based interpretation of the transgender experience risks leaving something out. A more robust reading, on which the subject experiences their body as, in some sense, not properly their own or not reflective of what they are, offers a straightforward explanation of why the distress takes this specifically embodied form. This does not show that the metaphysical reading is correct, but it does support the claim that it is a live and theoretically motivated option.

Finally, it is important to make explicit a methodological assumption guiding the present discussion. In philosophical contexts, claims about identity are often taken as one legitimate starting point unless there is a compelling reason to reinterpret them in weaker terms. This approach is reflected in influential work in the philosophy of language and metaphysics, where ordinary usage is often taken as a starting point for theorizing about reference and identity (Kripke 1980; Evans 1982; Lewis 1983). More recent work in analytic metaphysics and conceptual analysis similarly emphasizes the legitimacy of beginning from surface-level commitments embedded in ordinary discourse, absent strong defeaters (Thomasson 2015; Cappelen 2018). When a speaker asserts that they "do not match their body," one option is to treat this as a loose or metaphorical way of expressing facts about social role, presentation, or psychological discomfort. Another option, however, is to take the claim more literally, as expressing a tension between the self and the body in a way that bears on questions of personal identity. The present argument proceeds by exploring the consequences of adopting this latter interpretation. Unless there is decisive reason to rule it out in advance, it is methodologically legitimate to ask what follows if such claims are understood as making, or at least committing one to, a metaphysically robust assertion about the relation between self and body.

But, if this is how we understand the transgender experience, then either physiological continuity theories are incorrect, or claims about mismatch between self and body are incorrect. If identity tracks physiology (if physiological continuity theories of personal identity are correct), then when someone says that they do not match their body, what they are saying is not merely false but contradictory since, on this view, the predicate and subject cannot come apart in the relevant way without violating identity itself. This is a semantic claim about literal truth conditions under rigid designation, not a claim about psychological intelligibility or pragmatic usage. More generally, if a subject-term rigidly refers to X, then a denial of identity with X cannot be literally true under that semantics, since it would require X to both be and not be identical to itself in the relevant context.

At this point, a natural objection is that this reconstruction of the Popular Transgender Narrative is mistaken: one might argue that when transgender individuals say that they do not match their bodies, they are not making a claim about identity in the strict metaphysical sense, but rather about gender identity, social role, or lived experience. If that is right, then the tension identified here dissolves, since physiological continuity theories need not deny that there can be profound forms of psychological or social mismatch of this kind. However, this objection does not undermine the present argument so much as it redirects it. For the claim being made here is conditional: if the Popular Transgender Narrative is understood in the more metaphysically robust way outlined above, then the tension arises; if it is not, then an alternative reconstruction, of the sort considered below, will be more appropriate.

That is, the tension discussed here arises only if the Popular Transgender Narrative is read as asserting a mismatch in what one is, not merely in how one appears, is treated, or feels in contingent respects, and, again, an alternative approach will be explored below. For present purposes, it suffices to note that this is a philosophically demanding but defensible way of understanding such claims—and that, if physiological theories of personal identity are correct, then the sentence "I do not match my body" cannot be true on this interpretation because, on this view, I just am my body. Thus, either we must reject physiological continuity theories, or we must reject what I am calling the Popular Transgender Narrative: they cannot both be true under this interpretation. (The result is therefore conditional on adopting the metaphysically robust reading of the narrative outlined above.)

Proponents of psychological continuity theories of personal identity might welcome this result: they might hold on to the Popular Transgender Narrative while rejecting physiological continuity theories and never look back. Similarly, those who think that transgender individuals are laboring under an illusion also might welcome this result: they might hold on to physiological continuity theories while rejecting the Popular Transgender Narrative and never look back. However, in the remainder of this essay, I want to canvass two alternative ways of moving forward with the conjunction of the Popular Transgender Narrative and a physiological continuity theory, even while conceding that this requires some modification of one or the other.

The first way of moving forward, which is speculative and applies only in some contexts, argues that we should retain the Popular Transgender Narrative for pragmatic reasons, even though it is, strictly speaking, inaccurate. The idea is that, when there is a mismatch between body and mind of the kind experienced by transgender individuals, it is often easier to change, or at least to disguise, the mismatched aspect of the body (rather than the mismatched aspect of the mind).

For example, suppose that a trans woman believes that her body is incorrectly sexed. Her options include (i) trying to pass as a woman in public, (ii) undergoing gender reassignment surgery, and (iii) undergoing extensive psychotherapy in order to rid herself of the belief that her body is incorrectly sexed. Option (i) is typically a lower-cost solution: it might involve a change in wardrobe and grooming routine.[4] Options (ii) and (iii), by way of contrast, will be high-cost: they are financially expensive, bring risks of complications, and might not resolve the underlying issue. On this view, the options can be ranked, from best to worst, in the order in which they were enumerated: disguising the body

[4] Option (i) also might offer only temporary respite: it changes how much the individual thinks about the mismatch between body and mind, but the underlying mismatch remains.

(option (i)) is the first thing to try; then gender reassignment surgery (option (ii)); then psychotherapy (option (iii)).

These claims are illustrative. The ranking is intended only as a rough heuristic, not as a claim about all cases. However, if we accept this heuristic, then, according to this first way of moving forward, we can, and perhaps should, continue to talk about the transgender experience in terms of a mismatch between the self and physiology, rather than in terms of a mismatch between the self and psychology. That is, a proponent of this approach might argue that, although this way of speaking about the transgender experience is, strictly speaking, inaccurate (because physiological continuity is the correct criterion of personal identity, and physiological continuity is inconsistent with the Popular Transgender Narrative), so is our way of talking about the motion of the sun (e.g., we talk about the sun "rising" and "setting," as if it orbits the earth rather than the other way around).

Of course, the moral and social stakes are vastly different when it comes to conventions for talking about the sun and conventions for talking about the transgender experience. But the point for present purposes is merely the pragmatic retention of misleading language. Thus, someone who favors this first approach can contend that this way of talking about the transgender experience, while inaccurate in this strict sense, does no harm and, further, that there is a good reason for retaining it, namely: we know how to change and disguise the body and thereby allay the alienation inherent in the transgender experience in a way that we do not know how to change and disguise psychology. Thus, this approach, which depends on the empirical assumption that modifying or presenting the body is often more tractable than altering deep-seated psychological states, says that, although the Popular Transgender Narrative is, strictly speaking, incorrect (it is the psychology that does not match the self, not the other way around), we should retain it for pragmatic reasons.

An alternate way of moving forward is to reframe the transgender experience. This alternative is best understood as rejecting the metaphysical reading of the Popular Transgender Narrative rather than competing with it on its own terms. For example, Talia Mae Bettcher argues that transgender dysphoria does not arise from the belief that there is a mismatch between body and self, but, rather, from the belief that there is a mismatch between self and socially imposed roles and ways of being which are, according to Bettcher, incorrectly based on the body (Bettcher 2024). For example, on Bettcher's account, a trans woman might feel bad when she looks at her body, not because she thinks that it has the wrong sex, but because having her biological sex at the forefront of her consciousness reminds her of the painful fact that society will not allow her to compete as an athlete in the women's-only category, even though there is, according to Bettcher, no objective reason for this prohibition. Thus, according to Bettcher, a trans individual who maintains that their body has the wrong sex is using a sort of shorthand. It is not the sex of their body that is upsetting; what is upsetting is that society prevents them from doing something they want to do, and they are reminded of this latter fact when they see the sex of their body, because that is what the prevention is based on. But, if the distress is about social constraints rather than bodily misidentification, then there is no need to posit a divergence between self and body in the identity-constituting sense. This makes Bettcher's account of the transgender experience compatible with physiological continuity theories of personal identity, in the limited sense that it does not require denying them. In this way, Bettcher's account does not so much resolve the tension as dissolve it by rejecting the assumption that claims about transgender experience are claims about metaphysical identity in the first place.

Some might object that, because, on Bettcher's account, it is society that is in the wrong—according to Bettcher, the social rules that are in question should not be based on biological sex—transgender individuals should change society, not their bodies. However, Bettcher can respond with a cost-benefit analysis that mirrors the one from a few paragraphs up. That is, Bettcher might point out that it is much easier to change the body than society. Thus, Bettcher might conclude, while it might be good to strive for social change in the long run, for the concrete here and now, a transgender individual who experiences powerful and persistent dysphoria would do better to focus on what they can do to change (or disguise) their body.

Indeed, Bettcher may argue, further, that, because of the way in which human psychology works, transgender individuals might think that their body really is the problem, perhaps because they do not realize that the social roles and categories in which they want to participate are only incorrectly based on biological sex. In such cases, these individuals might say that their bodies do not match their selves. But, Bettcher might argue, these individuals need to dig deeper to get to the root of the problem, and, when they do so (Bettcher might say), they will see that their experience is consistent with physiological continuity theories, not because it supports them, but rather because it sidesteps issues about metaphysical identity altogether.

However, these two alternatives are not unassailable. For example, the first approach can be challenged by those who are sympathetic to the transgender community on the grounds that it makes transgender individuals into deceivers, and it also can be challenged by those who are unsympathetic to the transgender community on the basis of a Cliffordian commitment to the truth. Similarly, the second approach can be challenged on the grounds that some transgender individuals seem to assert that their dysphoria arises from the experience of their bodies directly, rather than from the fact that their bodies recall to them the painful exclusion from desired social practices—and some who are less sympathetic to the transgender community might challenge Bettcher's idea that these exclusionary practices should not be based on biology.

But the purpose of the present essay is not to settle these issues, nor is it to figure out what the best path forward is. My purpose is merely to highlight the tension between physiological continuity theories and the Popular Transgender Narrative, at least on one way of understanding it, and that aim has been accomplished.

References

American Psychiatric Association. 2022. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR)*. Washington: American Psychiatric Association Publishing. https://doi.org/10.1176/appi.books.9780890425787

Bettcher, T. 2024. *Beyond Personhood*. Minneapolis: University of Minnesota Press.

Bornstein, K. 2016. *Gender Outlaw*. New York: Vintage.

Britton, J. W., et al. 2016. "Electroencephalography." American Epilepsy Society. Available from: https://www.ncbi.nlm.nih.gov/books/NBK390356/

Cappelen, H. 2018. *Fixing Language*. Oxford: Oxford University Press.

DeGrazia, D. 1999. "Advance Directives, Dementia, and 'the Someone Else Problem'." *Bioethics* 13 (5): 373–391.

Erickson-Schroth, L. (ed.). 2022. *Trans Bodies, Trans Selves*. Oxford: Oxford University Press.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

Heinämaa, S. 2003. *Toward a Phenomenology of Sexual Difference*. London: Bloomsbury Publishing.

Kahn, S. 2014. *A Problem Based Introduction to Philosophy*. Dubuque: Kendall Hunt.

Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.

Lewis, D. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61 (4): 343–377.

McBee, T. 2014. *Man Alive*. San Francisco: City Lights Publishers.

Mock, J. 2014. *Redefining Realness*. New York: Atria Books.

Moschella, M. 2021. "Trapped in the Wrong Body?" *The Journal of Medicine and Philosophy* 46: 782–804.

Olson, E. 2024. "Personal Identity." In E. N. Zalta and U. Nodelman (eds.). *The Stanford Encyclopedia of Philosophy*, Winter 2024 Edition. URL = https://plato.stanford.edu/archives/win2024/entries/identity-personal/.

Pellegrino, E., et al. 2008. *Controversies in the Determination of Death*. Washington, DC: The President's Council on Bioethics.

Sauchelli, A. 2017. "The Animal, the Corpse, and the Remnant-Person." *Philosophical Studies* 174: 205–218.

Serano, J. 2007. *Whipping Girl*. Berkeley: Seal Press.

Skrzypek, J. 2023. "Personal Identity, Sexual Difference, and the Metaphysics of Gender." *Christian Bioethics* 29: 77–94.

Thomasson, A. 2015. *Ontology Made Easy*. Oxford: Oxford University Press.

Wiggins, D. 2003. *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.

World Professional Association for Transgender Health. 2022. *Standards of Care Version 8*. Abingdon: Taylor and Francis.

Croatian Journal of Philosophy
Vol. XXVI, No. 77, 2026
https://doi.org/10.52685/cjp.26.77.6
Received: January 23, 2026
Accepted: July 2, 2026

Self-Respect Is Incomplete Without Respect for Others

QING ZHAO
Southeast University, Nanjing, PR China

In Kantian ethics, self-respect is a fundamental duty to oneself, embodying the commitment to uphold one's dignity as a moral agent. This commitment, which proscribes self-degradation, is inherently structured by both intrapersonal and interpersonal dimensions. This paper argues that because the exercise of self-respect occurs primarily in the social sphere, a self-respecting person's regard for her own dignity is inseparable from how she relates to others. By analyzing how vices of disrespect toward others such as arrogance, defamation, and ridicule constitute the opposite of self-respect, and by developing an intersubjective account of dignity, this paper demonstrates that respect for others is a constitutive part of the Kantian duty of self-respect.

Keywords: Self-respect; dignity; servility; arrogance; sympathy

Introduction

In the *Metaphysics of Morals*,[1] Kant characterizes self-respect (*Selbstachtung*) as a duty to uphold the dignity of humanity[2] within us (MS

[1] I use the following abbreviations: *GMS*= *Groundwork of the Metaphysics of Morals* (in Kant 2011); *MS* = *The Metaphysics of Morals* (in Kant 1991); *LE*= *Lecture on Ethics* (in Kant 1997); *RGV* = *Religion within the Boundaries of Mere Reason* (in Kant 1998); *Päd* = *Lectures on Pedagogy* (in Kant 2007); *AL-F*=*Lecture of the Winter Semester 1775–1776* (in Kant 2013); *Anth* = *Anthropology from a Pragmatic Point of View* (in Kant 2007); *KpV*= *Critique of Practical Reason* (in Kant 2015).

[2] While Kant distinguishes the predisposition to humanity from personality (RGV 6:27-28), he holds that humanity possesses dignity by virtue of its capacity for morality (GMS 4:435), and, in discussing duties of virtue, he speaks of "one's intelligible being (moral disposition)" and the dignity of humanity (MS 6:436, 462). I agree with Timmons (2017: 220n) that Kant employs *humanity* in a broad sense that includes personality in the *Doctrine of Virtue*. Thus, to respect the dignity of humanity is to value the potential moral capacity inherent in all human beings. For

6:436). Notably, the distinction between *self-respect* and *self-esteem* is not always clearly maintained in Kant's texts. In this paper, I prefer the term *self-respect* because it better captures the idea of respecting one's potential capacity for morality as such. To respect oneself, Kant tells us, is to be a self-respecting person: someone who recognizes and affirms her own dignity, and who therefore is not servile, does not ignore her equal status and moral rights when compared with others.

A prevailing view in the secondary literature holds that the duty of self-respect is fundamentally a duty to oneself.[3] On this understanding, self-respect is not primarily other-regarding, nor can it be reduced to a merely prudential concern for one's own well-being. Rather, it belongs directly to the agent's own moral perfection.[4] But can self-respect truly be realized independently of respect for others? In this article, I argue that the common reading overlooks the social dimension of self-respect and its connection to our duties to others. To bring this dimension to light requires a reconception[5] of self-respect that reveals its inherently social structure. Kant's text suggests a necessary connection between self-respect and respect for others (MS 6:462). I argue that this connection is constitutive: respect for others is internally related to the very concept of self-respect. In other words, the duty of self-respect already contains within itself the duty to respect others.

People who respect themselves, I argue, cannot fail to respect others: they won't focus exclusively on their own dignity, and they won't act arrogantly toward others. If we want to live up to the dignity of humanity, it would also seem to consist in feeling sympathy and anger when witnessing others being humiliated. Indeed, since respect for others is *constitutive of* self-respect, the duty of self-respect cannot be properly understood in isolation. I will therefore begin by analyzing the demands that self-respect places on the individual, and then show how this duty already contains, within its very structure, the requirement to respect others.

I proceed in five sections. I begin by elucidating the dual interpersonal and intrapersonal structure of Kantian self-respect. In the second section, I turn to the vices of arrogance, defamation, and ridicule. By examining how these vices are incompatible with self-respect, I offer a

debates on Kantian humanity and dignity, see Wood (1999: 118–121); Dean (2006); Møller (2024: 45–58); Willaschek (2024: 30).

[3] The distinction between duties *to* and duties *regarding* is significant but will not be examined in full here. For present purposes, to say that self-respect is a duty to oneself is to say that it pertains fundamentally to one's own humanity as a moral agent. For discussions of the duty *to/regarding* distinction, see Timmermann (2006) and Johnson (2010).

[4] See Timmermann (2006), Guyer (2010), Denis (1997, 2015), Massey (1983), Hill (1973, 2021), Wood (2008), and Timmons (2017).

[5] I am grateful to the anonymous reviewer for clarifying the distinction between the extension reading and the reconception reading of self-respect, and for pointing out a confusion between these two approaches in my earlier argument. Those comments have helped me to sharpen the core thesis of this article considerably.

negative argument: to disrespect others is to disrespect oneself. This establishes that self-respect cannot coexist with disrespecting others. However, this negative claim alone does not yet tell us why self-respect positively requires respecting others. The third section then builds the positive argument. Drawing on the intersubjective structure of dignity—that the dignity I respect in myself is the same dignity I must respect in all others—I demonstrate that respect for others is not merely compatible with self-respect but constitutive of it. In other words, self-respect necessarily includes respect for others, not as an external derivation but as an internal requirement. The fourth section addresses a potential objection: that the duty of self-respect is too demanding, and that it would therefore lead us to condemn those who suffer moral failure. In response, I argue that self-respect, properly understood, does not demand such condemnation. On the contrary, its inherent requirement to value the rights of all persons compels a necessary sympathy for those whose rights are trampled on yet who cannot resist. Finally, I address two possible objections to my argument: first, that the real world's complexity compels us to compromise Kantian self-respect, and second, that sympathy for the servile risks *rationalizing* their moral failure.

1. Kant's definition of self-respect

In the *Metaphysics of Morals*, Kant holds that all human beings equally possess dignity by virtue of their humanity. This dignity directly grounds the duty of self-respect. Such a duty consists in treating oneself as the equal of every other person (MS 6:435) and thus requires one to uphold one's dignity rather than demean oneself for personal benefit. The feeling of noble pride is not the ground of this duty but rather a natural accompaniment that arises when we recognize ourselves as fulfilling it. "Humanity in his person," as Kant puts it, "is the object of the respect which he can demand from every other man, but which he must also not forfeit" (MS 6:435), and if we fail to defend our dignity, this means we choose to forfeit our dignity and instead make ourselves into a thing,[6] which is servility.

Servility, by contrast, is an attitude whereby one believes that one is inferior to others,[7] "waiving any claim to moral worth in oneself, in the belief that one will thereby acquire a borrowed worth" (MS 6:435). As a result, one also waives the right to demand respect from others and seems thereby to tacitly allow others to insult and demean one, thus regarding the violation of one's human dignity by others as mor-

[6] Bunch (2014: 82) holds that if we cannot claim the respect that dignity authorizes us to claim, this means we throw away our humanity and become a thing.

[7] Some scholars, Timmons (2021: 179) and Stark (2022: 271) share this view. In contrast, Massey (1983: 65) understands servility as behavior motivated by desire rather than by a belief in one's moral inferiority.

ally acceptable.[8] However, a servile person's subjective waiving of their claim to respect does not in fact relieve others of their duty to respect that person's dignity.[9] For dignity, as an inalienable right possessed by everyone, cannot be waived. There is thus an asymmetry here: an invitation to be mistreated does not give others a right to mistreat.

In his influential work, Thomas E. Hill, Jr. (1973) provides three vivid illustrations of how individuals fail to recognize their own dignity and rights: (1) Uncle Tom, who believes that Black people cannot possess the same rights as white people; (2) the self-deprecating man, who thinks he must earn his rights through hard work; and (3) the deferential wife who regards herself as an obedient wife who serves her husband.[10]

According to Hill, human beings have basic moral rights, and each person's recognition of his moral rights is a reflection of his self-respect. Servility, the opposite of self-respect, entails a failure to properly acknowledge or uphold one's moral rights. This comes in two forms: "one resulting from misunderstanding of one's rights and the other from placing a comparatively low value on them. In either case, servility manifests the absence of a certain kind of self-respect. The respect which is missing is not respect for one's merits but respect for one's rights" (Hill 1973: 97).

But is the moral failing of servility merely a failure to recognize one's moral rights? The answer is clearly no. Servility is reprehensible not simply because we are misunderstanding our rights, but because we engage in a form of *moral self-repudiation*, an act of *denying our own dignity* from a moral standpoint, constituting a profound self-abasement or "the deliberate and so culpable devaluing of oneself" (Dillon 2004: 200). Consequently, the most fundamental demand of the duty of self-respect is the *affirmation of one's own worth*. (Obviously, there is a further question about whether misusing one's body or humanity, through acts such as gluttony or lying, entails a loss of self-respect, although we can set aside that issue for present purposes.)

In his remarks on how to teach children to perform the duty to oneself, Kant suggests that one can violate one's dignity either by drinking or by other unnatural vices that degrade oneself below the animals, or by groveling before others (Päd 9:489). I take it that Kant's point is that the duty of self-respect lies in acknowledging the inherent dignity of one's rational nature and acting in ways that affirm this value.[11] From this standpoint, it might seem that any immoral behavior, regardless

[8] Self-respect represents our moral attitude toward those who insult or harm us, such as "how I ought to let others treat me" (Bacin 2013: 263).

[9] I am grateful to the reviewer for this suggestion.

[10] This influential view has been considered and challenged by some scholars, see Boxill and Boxill (2015).

[11] Denis (2015: 215) offers a similar view, holding that the objective principle of self-esteem consists in acting in accordance with the dignity of humanity.

of how minor, can be seen as a failure of self-respect,[12] insofar as it involves treating one's humanity as a mere means, which is "destroying ourselves merely to achieve contingent ends" (Dean 2006: 138).

Thus, as a duty to oneself, self-respect is a duty to treat our humanity as an end that has both *interpersonal* and *intrapersonal* aspects. In its interpersonal sense, self-respect is a duty that is characteristic in its object (it is a response to one's inherent dignity) and in its content (it avoids servility or false humility, affirming *my* equal standing with others). Used in its intrapersonal sense, self-respect refers to the requirement of safeguarding the dignity of one's humanity.[13] Clarifying this point is crucial for a proper understanding of the duty of self-respect. It prevents the conflation of all failures of self-respect with servility, and resists the temptation to use servility as a blanket explanation for all phenomena of self-disrespect.

To this point, I have approached the duty of self-respect from the standpoint of how one treats oneself. I now turn to the duty of respecting others: in what sense does respecting others constitute part of the duty of self-respect? Before addressing this question, I must first examine how the vices that involve disrespecting others are incompatible with the duty of self-respect.

2. Disrespecting others, disrespecting oneself

I have argued that the affirmation of one's own dignity required by the duty of self-respect is determined in relation to others, namely, seeing oneself as an equal among equals. This section will further argue that a person who disrespects others, for example, by treating them arrogantly, cannot be a person with genuine self-respect. I will then show, one by one, why the vices of arrogance, defamation, and ridicule are fundamentally incompatible with self-respect.[14]

Arrogance. It is commonly held that arrogance toward others is an excess of self-respect, and servility a deficiency, with true self-respect being the mean between them. This section rejects that picture. I argue that the arrogant person lacks self-respect. Arrogance towards others often manifests in interpersonal interaction through an exaggerated sense of self-value and an expectation that others will bow and scrape. Kant saw this arrogance as a kind of madness or foolishness (MS 6:466; AL-F 25:1306; LE 27:458; Anth 7:217), and it leads people not to a genuine sense of pride in their dignity, but to "strive after *the*

[12] Massey (1983: 67–68) contends that all immoral acts entail a lack of self-respect, because they involve failing to treat oneself in a manner befitting one's moral capacity.

[13] Kantian self-respect entails not only Darwall's recognition self-respect (respect for humanity's dignity as a constraint) but also an appraisal self-respect that requires making oneself worthy of honor through *merit* (LE 27: 665). For the distinction between two kinds of self-respect, see Darwall (1977: 47–49).

[14] I am very grateful to the reviewer for asking me to clarify this point.

reputation of honor" (Anth 7:273), a *passion*—what he called "*Mania for honor*" (Anth 7:273). An act of arrogance, which constitutes a fundamental denial of others' humanity because it "demands from others a respect it denies them" (MS 6:466), not only injures their self-respect[15] but also constitutes a failure of self-respect in the arrogant individual.

An arrogant individual seeks to bolster their sense of superiority by belittling others, which is precisely a marker of deficient self-respect because they not only "freely follow his inclinations" (LE 27:345) but also seek validation through comparisons or external approval to feel confident in their own worth. In the *Lecture on Ethics*, Kant asserts that when we command others to honor us or when we think less of others, we run the risk that the other people will hold us in contempt. The idea behind this argument is that there is a paradox in the arrogant individual:

> Such arrogance always betrays little understanding, for it is the act of a fool not to strive for the possession of true worth by self-mastery and the trustworthy performance of duty on principle but, rather, to seek honour in casual outer circumstances, paying no respect to one's humanity, and thus demeaning himself (LE 27:667).

The arrogant individuals, who project an image of superiority, believe themselves to be above others. In reality, however, they are deeply dependent on external validation, which ultimately reveals a fundamental lack of confidence in their own intrinsic worth. As Dillon rightly points out, "The deeper problem is that he cannot regard any being as an end in itself, as unconditionally deserving of respect, himself included" (2004: 201). In this meaningless pursuit, they *enslave* themselves to the opinions of others, becoming prisoners of their own making. Moreover, by equating their value with these elusive external accolades, they have, in essence, abdicated the very claim to their human dignity. Therefore, while we might describe arrogance as an excess of self-respect, it is more accurately understood as a form of its failure. That is to say, arrogant persons lack self-respect because they fail to recognize that they possess a human dignity that is certain and requires no external praise.

Defamation. It is commonly thought that the disrespectful act of spreading rumors about others harms their self-respect, because it deprives them of what Fahmy (2013) calls moral privacy. However, I argue that when a person defames another, she also shows disrespect

[15] One might argue that a person with genuine self-respect would be immune to such arrogance. However, this overlooks a fundamental human nature: we are social beings, inherently vulnerable to the judgments of those around us. This is especially true in intimate relationships or situations of power imbalance, where others' condescension and dismissal can make even the most secure individual question their own worth: Am I really worth nothing? A particularly evident example is *gaslighting*: after enduring long-term suppression and denial from others, the victim will inevitably begin to doubt their own abilities to make judgments, and even the validity of their emotional needs. See Stark (2022).

toward herself. This can be seen by examining the structure of the defamatory act itself:

(1) When S defames P with respect to some fact F, S makes a judgment that F is a morally defective state of affairs, and that F will arouse negative moral judgments about P in others.
(2) S does not discuss F openly in P's presence. Moreover, S's motive is neither to protect the audience from harm by P, nor to promote P's moral improvement.
(3) Rather, S's aim is simply to damage P's reputation.

Whether F is a flaw invented by S or a real defect of P, it should not be a reason for P to be disrespected by others. By spreading P's defect and inviting others to join in judging P, S shows an intention to deprive P of the respect to which P is entitled. Thus, following Aquinas (1920: II-II, q. 73), we might understand defamation as a form of *secret theft*, namely the theft of the respect that another deserves. It is precisely in this secret theft that the defamer loses her own self-respect.

First, (1) shows that the defamer possesses the capacity for rational judgment: she can distinguish right from wrong. Yet (2) and (3) show that she freely chooses to do wrong. The defamer therefore actively chooses to degrade the humanity in her own person—the humanity that is essential to human nature—into a mere tool for harming others to satisfy her own inclinations. Or, in Kant's terms, her person is governed by her inclinations, and she thereby makes herself "lower even than the animals" (LE 27: 345). In this sense, she fails to meet the demands of self-respect.

Second, the secretive nature of defamation reveals precisely that the defamer lacks an inner sense of her own worth. In other words, she is unaware of her own dignity. That is why she dares not compete openly, nor confront directly, but can only fill the void of her inner worth by stealing the honor of others, "since in that case our own faults seem small" (LE 27: 458). Hence, the defamer not only externally deprives others of the dignity and honor they deserve; through her very act, she also displays the fact that she is unaware of her own dignity. This is, without question, a failure of self-respect.

Ridicule. Ridicule makes others' pain or mistakes the object of one's own amusement. Unlike defamation, which is conducted behind the back of its victim, ridicule occurs in the presence of the person ridiculed. For precisely this reason, it damages the victim's self-esteem far more severely than calumny does, since the victim is forced to hear these malicious jokes firsthand. Ridicule thereby signals to its target that one has forfeited any claim to respect by virtue of one's own failings, and in so doing, it wholly denies the victim the moral standing to be taken seriously as a member of the moral community (Fahmy 2013). If calumny is a secret theft of another's dignity, then ridicule is its public counterpart. In ridiculing others, the agent treats her own humanity merely as a means of gratifying her malicious inclinations.

Moreover, the scoffer thereby reveals her true moral character: she has no grasp of the dignity of humanity. It is precisely for this reason that she needs to ridicule others in order to satisfy her sense of superiority. Therefore, the ridiculer's sense of self-worth is inherently unstable: she requires both a victim and an audience to sustain it. When there is no one left to mock, her self-worth simply evaporates. This dependence on external conditions reveals precisely that she lacks genuine self-respect. A person with true self-worth does not rest it on the failures or embarrassments of others; she need not ridicule anyone to assure herself that she is good, because she has already grounded her value on a basis that requires no comparison.

From this analysis of arrogance, defamation, and ridicule, it becomes clear that those who fail to respect others are also lacking in self-respect. This is evident in two respects. First, arrogance, defamation, and ridicule share a common structure: each seeks to ground one's self-worth in the negation of another's dignity. In arrogance, the agent elevates herself by refusing to acknowledge others as equals. In calumny, she feeds her vanity by tarnishing another's reputation. In ridicule, she derives a sense of superiority from making others appear foolish. What unites these attitudes is that the agent's self-worth is fundamentally comparative and depends on the perceived inferiority of others. This reveals that such agents fail to recognize the inherent dignity of their own humanity. In short, they lack genuine self-respect.

Second, by following their own inclinations and allowing self-love to become the principle of legislation, they fall into a state of self-conceit, a condition that stands in clear opposition to genuine self-respect. Consider the following passage in the *Critique of Practical Reason*:

> But it *strikes down* self-conceit altogether, since all claims to *esteem for oneself* that precede accord with the moral law are null and quite unwarranted because certainty of a disposition in accord with this law is the first condition of any worth of a person, and any presumption prior to this is false and opposed to the law. Now, the propensity to *self-esteem*, so long as it rests only on sensibility, belongs with the inclinations which the moral law infringes upon. (KpV 5:73; emphasis mine)

This passage tells us that the legitimacy of one's claim to dignity, such as declaring that one respects oneself, depends entirely on whether one has first satisfied the demands of the moral law, for example, the duty to respect others. Any presupposition of selfrespect that precedes this is "false and opposed to the law". It is only by acting in accordance with the moral law, and thereby living up to the demands of human dignity, that we are entitled to claim self-respect.[16]

[16] It is important to note that this is not to say that selfrespect *simpliciter* is equivalent to obeying the moral law. There is indeed a sense in which selfrespect is equivalent to (or follows from) the respect for the moral law within us, and in this sense, it is a feature that all rational beings, as such, unavoidably possess. But this is not the sense of selfrespect with which the present paper is primarily concerned. If selfrespect were understood in this way, then the claim that selfrespect includes

Arrogance, backbiting, and ridicule, which are vices of disrespecting others, first presuppose one's own supreme and superior status so that they can grant themselves the right to demean and mock others, thereby refusing to fulfill the duty of respect one owes to them. At the core of such conduct lies a form of self-conceit in which the agent takes the satisfaction of her own sensible inclinations as the principle of legislation and pretends that this inclination can license her to demean others, a point on which Russell (2020) argues that self-conceit involves a fantasy of dictatorial authority. Kant makes it clear that this attitude of living "without any rule, merely following the inclination one has devised for oneself" makes a person "the most abhorrent of objects" (LE 27:346). Consequently, disrespecting others is not merely a denigration of them but also a denigration of the agent herself because, through self-conceit, it becomes difficult for her to meet the demands of human dignity.

I have now shown that arrogance, defamation, and ridicule are not merely violations of duties to others; they are themselves manifestations of a failure of selfrespect. A person with genuine selfrespect could not engage in such conduct. However, this section has only completed a negative argument: a person with selfrespect does not do X (i.e., disrespect others). What remains to be shown is that a person with selfrespect does do Y (i.e., actively respects others), that is, that respecting others is itself a constitutive element of selfrespect. The next section will develop this positive argument.

3. Self-respect as relational

On the issue of self-respect and respect for persons, Adam Cureton approaches this from the perspective of *reciprocity,* and he suggests that the duty of self-respect leads to respect for others because we need to respect others in order to gain their respect, thereby protecting our own self-respect. As he says,

> We are rationally disposed to pursue our own happiness and respect ourselves, but also to give the same help and respect to others in return for their help and respect, so duties of beneficence and respect for one another are justifiable to everyone in virtue of our self-regarding rational concerns. (Cureton 2013: 182)

In short, Cureton's view is that I respect you because I was hoping you could respect me, as self-respect is very important to me. If I un-

respect for others would amount to nothing more than an analytic proposition, carrying no independent argumentative weight. This paper is concerned with a different sense of selfrespect: namely, the individual's duty to live in a manner befitting her own moral agency, and to express in her actions and attitudes an appropriate regard for her own dignity. Understood in this sense, selfrespect is not something that everyone automatically possesses; rather, it requires active cultivation. Hill (2021: 238) offers a detailed explanation of these two kinds of respect.

derstand him correctly, Cureton's argument rests on a condition of reciprocity: we respect others in the expectation that they will reciprocate, thereby helping us to achieve self-respect. This renders respect a conditional and instrumental act of exchange. If the other person is entirely incapable of returning respect, a rational agent would have no reason to respect that person. This position, however, conflicts with Kantian ethics. Consider a patient in a vegetative state. Within Cureton's framework, no possible return is available; therefore, a rational agent would have no reason to respect such a patient. Yet for Kant, the objects of respect are all rational beings, regardless of whether they can reciprocate. We have a duty to respect a vegetative patient not because his respect could bring me anything, but because he himself possesses a dignity that deserves respect. Thus, Cureton's reciprocity model cannot account for the unconditionality of the duty to respect persons.

Thus, I contend that any reciprocitybased approach to the relationship between selfrespect and respect for others is misguided. Respect for others is not instrumentally related to selfrespect but constitutively so; that is, respecting others forms part of what it means to respect oneself. To clarify this point, we must reconsider the structure of the concept of dignity.

When a person refuses to be servile and insists on her own dignity, what exactly is she asserting? She is asserting that, as a rational being, she possesses a noncomparative and absolute value that cannot be diminished by others. The key point is that this assertion is not an exclusive one. To claim "I have dignity" is not to claim that "only I have dignity". On the contrary, the very concept of dignity carries with it a universal commitment. If I have dignity, then every rational being of my kind has the same dignity.

It can be seen that dignity itself possesses an *intersubjective* structure. That is to say, the concept of dignity already presupposes a relation between *you and me*. As Kant puts it, when my selfrespect gives me the right to demand respect from others, I ought also to take on the duty to respect them in return (MS 6:462). Hence, to respect the dignity of one's own humanity implies respecting the dignity of the humanity in others. This is not an externally imposed requirement. Rather, it follows from the fact that the concept of dignity itself admits of no exceptions. It is in this sense that selfrespect already contains respect for others. This is what I mean by a *reconception* of selfrespect. Respecting others is not an additional requirement that extends outward from selfrespect. It is instead a constitutive part of selfrespect itself. The answer becomes clear once we note that for Kant, the nature of respect for *my humanity* is identical to respect for *everyone's humanity*. In the *Metaphysics of Morals*, Kant says:

> Every man has a legitimate claim to respect from his fellow men and is *in turn* bound to respect every other …. But just as he cannot give him-

> self away for any price (this would conflict with his duty of self-esteem), so *neither can he act contrary to the equally necessary self-esteem of others*, as men, that is, he is under obligation to acknowledge, in a practical way, the dignity of humanity in every other man. (MS 6:462; emphasis mine)

What is truly required by self-respect cannot be separate from respecting others. Sometimes Kant says self-respect and respect for others are mutually related; sometimes he suggests that self-respect requires us to protect others' self-respect: in either case, self-respect requires us to respect others.[17] If our duty to respect ourselves is constructed so as to enable us to recognize that we share an equal moral status with all others, then respecting others is not an additional requirement derived from self-respect, but rather a constitutive element of self-respect itself. For to anyone who would claim that we can respect our own humanity while denying others, we can always ask: "Do we truly respect the dignity that is owed to everyone?" Since the humanity we respect in ourselves is the very same humanity shared by all rational beings, then if a person fails to respect others, he is not someone who has self-respect but has not yet extended it outward; rather, he could never have possessed genuine self-respect in the first place. Failing to do so would amount to a contradiction in our moral attitude toward humanity itself.

Timmermann (2006: 508), however, invokes the vivid metaphor of "desert island duties" to explain duties to the self. In his view, the ground of duties to the self does not depend on moral interaction with others, nor does it vary with the presence or absence of others. Understood in this way, self-respect, as a duty to oneself, seems to concern only one's own rational nature, while relations with others are merely contingent and external consequences.

Nevertheless, the intersubjectivity of dignity implies that even if the ground of the duty of self-respect does not arise from interaction with others, our understanding and *fulfillment* of this duty still intrinsically involve relations with others. "Desert island duties" successfully rules out the mistaken view that the duty of self-respect originates from the contingent needs of others, yet this should not be misunderstood as claiming that the duty of self-respect can be realized in complete isolation from others. A person who lives forever on a desert island, though still under a duty not to commit suicide or not to treat himself as a mere instrument, would fail to understand why servility is a form of self-harm, for this moral phenomenon itself presupposes the presence of others. Thus, the desert island metaphor reveals the ground of duties, but it does not deny the social dimension of self-respect mandated by the intersubjectivity of dignity. A person who has self-respect must also respect others. This must express a necessity: it is not the case

[17] A similar view is briefly suggested by Wood (2008: 174), who holds that, in relation to others, our duty of self-respect is twofold: we must avoid both arrogance and a servile disposition.

that someone with self-respect will more easily respect others, as Denis (1997: 343) suggests, that one's attitude toward oneself affects how one treats others. Rather, respecting others is a necessary condition of being a person with self-respect.

So far, I have reconceived the duty of self-respect without blurring the distinction between duties to oneself and to others. Both duties share the same ground—the dignity of humanity—whose intersubjectivity entails that recognizing one's own dignity just is recognizing the dignity of others. Yet they remain distinguishable by their objects: duties to oneself are directed at one's own rational nature, and duties to others at the rational nature of others.

Given that self-respect not only concerns oneself but also intrinsically demands respect for the personality and dignity of others, a question follows: how should we treat those who fail in their duty of self-respect? Common discourse tends to condemn such moral failures, assuming that the lack of self-respect results solely from personal weakness or choice, for which the individual alone is to blame. This view, however, ignores the significant role of social circumstances in shaping self-respect. Accordingly, I will argue that a genuinely self-respecting person ought not to despise those without self-respect, but instead to regard them with sympathy. It is in this sense that the duty of self-respect is a duty we owe to all persons.

4. Self-respect with sympathy

In this section, I argue for a deeper social implication of Kant's duty of self-respect: precisely because this duty itself incorporates a regard for the dignity of others, this regard elicits sympathy, which in turn precludes the blaming of those who fail in self-respect. Consider the case of servility as a failure of self-respect.[18] First, we must first clarify what opposing servility means for the self. The duty of self-respect suggests that we are entitled to *feel* noble pride in our humanity and *demand* respect from others,[19] which can prevent people from being treated as mere means.[20] When others demand that we debase ourselves, the duty of self-respect requires us to firmly believe in our equal moral standing with anyone. In the face of humiliation, our innate sense of dignity will inevitably evoke moral anger. More importantly, while we may pursue our own happiness and purposes, we must never sacrifice our dignity for external gains. Put simply, the duty of self-respect requires us to stand up for our claim to respect before others.

[18] For reasons of space, I focus here on servility. But the discussion is meant to apply also to other cases of failures of self-respect.

[19] This idea is inspired by Fahmy's (2013) view that we can understand the respect and dignity entitled to us by asking what vices deprive us of.

[20] Boxill (1995) argues that protesting injustice expresses self-respect. I follow this view, understanding servility as the acceptance of inferiority, which undermines respect for persons.

Admittedly, in certain situations, servility may appear in a *cunning* form: Either I might be fully aware of my own dignity, yet choose to act obsequiously for personal interests, or I might be completely unaware that such behaviors constitute servility and will inflict harm upon my self-respect. However, whether one merely performs servility or genuinely believes oneself to be inferior, or even acts out of sheer ignorance, the very moment they place their own dignity beneath the feet of others constitutes servility, a fundamental violation of one's duty to self-respect.[21] As Kant puts it, "bowing and scraping before a man seems *in any case* to be unworthy of a man" (MS 6:437; emphasis mine).

Kant's response may seem strikingly austere and is open to criticism, as it appears to neglect the fact that servility, whether conscious or out of ignorance, is often a product of social oppression. Erica Holberg argues that Kant overlooks a critical reality: in certain contexts, the inability to maintain or even recognize self-respect stems not from individual defect but from systemic structures. As she says, "The concern is that insofar as Kantianism fails to acknowledge the ways immersion in oppressive social environments can warp an individual's understanding of what she is owed and capable of as a moral agent, it perpetuates the cruelty of unjust social systems in the guise of respecting individual autonomy" (Holberg 2017: 304).

If it is obligatory for every individual, then a failure to uphold this duty constitutes a moral failing, rendering the person liable to reproach—such as being asked, "Why do you demean yourself?" or "Why do you not resist?"—implying that they ought to have done better. Moreover, placing the full burden of action on the oppressed neglects the significant sacrifices often entailed in acts of resistance. For instance, refusing to ingratiate oneself with a colleague or superior may lead to the loss of steady employment, potentially exacerbating the individual's precarity. It is in this context that scholars such as Serene Khader suggest that concerns regarding self-respect should not be decisive in judging whether agents ought to comply with oppressive norms. What matters, she contends (2021: 240), is that the oppressed individual continues to regard themselves as equal and the system as unjust, a perception that, in itself, performatively constitutes an act of resistance.

From these debates, it becomes evident that the obligation to respect oneself is not fundamentally disputed. Rather, the crucial questions become: What does it truly mean to respect oneself under structurally unjust conditions? Is the failure to resist the individual's responsibility? And are we justified in condemning those who are unable to rebel?

While I have considerable sympathy for their concerns, I must concede that the Kantian duty of self-respect does indeed offer a stringent answer to the first two questions. Although Kant does not explicitly

[21] As Hay (2011: 39) contends, while ignorance may excuse a servile person from *blame*, it does not absolve them of the *obligation* to resist oppression.

address systemic injustice in his discussion of this duty, we can infer from his remarks— "Be no man's lackey. Do not let others tread with impunity on your rights. Contract no debt for which you cannot give full security. Do not accept favors you could do without, and do not be a parasite or a flatterer or (what really differs from these only in degree) a beggar" (MS 6:436-437)—that servile behavior is precisely directed toward those in positions of social dominance, such as the wealthy who control resources or the powerful who can violate others' rights with impunity.

If this reasoning holds, we may conclude that the Kantian conception of self-respect demands that even within an oppressive social context, when facing dominant parties like the powerful or the wealthy—though Kant does not explicitly call for resisting oppression, which distinguishes his view from Hay's (2011) interpretation—we must express this rejection *externally* by refusing to submit, tolerating no humiliation, and avoiding indebtedness. If we nevertheless act in ways contrary to our own dignity, Kant would hold us fully responsible for such moral failure. As he starkly puts it, "Complaining whining, even crying out in bodily pain, is unworthy of you, especially if you are aware of having deserved it" (MS 6:436-437).

It is reassuring that, despite Kant's stringent demands regarding duties to oneself, this does not imply that he believed individuals failing to act self-respectfully deserve blame from others.[22] Kant's discussion of sympathy helps clarify why a person can be responsible for their servile actions yet not be blameworthy, affecting a clear separation between responsibility and external blame.[23]

Kant distinguished between two objects that evoke human sympathy: the suffering of others and the injustice they endure. The former leads us to feel sorry for their misfortune, a kind of sympathy that can be understood as passive, in which one merely involuntarily feels the suffering of another; however, the anger aroused by witnessing others being humiliated or having their rights violated is a *noble* form of sympathy (AL-F 25:606), and this constitutes active sympathy. This active sympathy proves that we deeply cherish and recognize the absolute value of rights themselves. Indifference towards a victim of insult indicates a fundamental disregard for the rights of humanity. As Kant puts it, "those who do not have such sympathy, do not highly value the right of other human beings." (AL-F 25:606)

Now, let us apply this argument to the problem of victim-blaming. Suppose that there is a person who is trapped in a servile relationship with their supervisor, submitting to every unreasonable demand and

[22] Baron (1985: 393) holds that wrongness is distinct from blameworthiness.

[23] While Blöser (2015) has worked to separate responsibility from blameworthiness, her argument fails to persuade Holberg (2017: 301), who maintains that Kantian responsibility is inseparable from moral praise or blame. This paper addresses Holberg's concern by treating Kantian sympathy as a practical resource, deliberately bypassing its metaphysical foundations.

enduring verbal abuse without resistance. What we should do for that person is to feel anger at the injustice they face and extend care and understanding by acknowledging their struggles, helplessness, and pain from their own perspective. Suppose you once lived in servility yourself but are now striving to fulfill Kant's duty of self-respect. In such cases, we ought to respond with sympathy toward this person and feel more keenly their emotional and moral anguish—not blame them for failing to resist their supervisor. This is because, according to Kant's sympathy, we need to place ourselves in others' positions to genuinely grasp their suffering. Therefore, as individuals who have struggled against servility ourselves, we are better able to understand the difficulty, sacrifice, and courage required to reject such submission. If sympathy not only allows us to better understand, even communicate, the pain of wounded self-respect, but also enables us to see the other side of their story: the hidden struggles behind their choices, then we cannot adopt a condescending attitude to blame those who fail to fulfill the duty of self-respect.

It follows that sympathy for others reflects a regard for their dignity, and this regard is precisely the moral concern of a person with self-respect: I value not only my own dignity but also yours. Even if another has voluntarily forfeited his or her dignity, I still have an obligation to respect and defend it. By contrast, if a servile person has internalized an unjust system of oppression, her lack of self-respect will inevitably lead her to demean others who are in the same oppressed situation; she will fail to recognize the dignity of others.[24] Thus, a person who strives to defend her own self-respect under oppressive conditions will recognize the struggle involved in upholding human dignity. She will also recognize that others in the same situation who lack self-respect are the result of an unjust order that distorts human moral connection. Although she may regard these failures of self-respect as morally wrong, she does not condemn or entirely reject the person herself. Instead, our compassionate response and our insistence that "this is not how the world should be" is an expression of fundamental respect for their personhood. It affirms our belief that, beneath the layers of oppression, they retain the potential capacity for morality and deserve our respect.

In my view, the dignity of each person is interconnected. Our commitment to upholding our own dignity necessarily involves a concern for the dignity of others. This is the obligation that a self-respecting individual must strive to fulfil. Only by respecting the dignity shared by all can self-respect be fully understood as a concern that is not merely self-regarding but also other-regarding, thereby satisfying the call for care toward all humanity that Kant issues at the conclusion of his *Lectures on Pedagogy*: "In our soul there is something that makes us take an interest 1) in our own self, 2) in others with whom we have grown up, and then also 3) an interest in the best for the world must come to pass" (Päd 9:498).

[24] I would like to thank the anonymous reviewer for raising this point.

5. Objections

Perhaps the most obvious objection to a stringent duty of self-respect is that it fails to account for necessary compromises under duress. As established, Kant's duty of self-respect is stringent. While the theory of sympathy may prevent victim-blaming, this does not address a more fundamental question: can Kantian self-respect admit any compromise?

I argue that temporary compromise is permissible only as a provisional strategy that must be directed toward future action. This posture involves, first, a firm inner conviction during periods of compliance: one must consciously affirm that the situation is unjust and that their compromise is a strategic sacrifice, not a moral surrender. Second, true self-respect demands action; it inherently motivates us to alter or flee unjust structures when possible.[25]

Let me now turn to a second possible objection, which concerns whether sympathizing with servile individuals risks *rationalizing* their immoral conduct. Experiences from daily life give us reason to think that when confronted with substantial temptations of benefit, most people have already decided to set aside their personal dignity. We tend to believe that those who obtain great benefits through various servile acts and subsequently indulge in hedonistic pleasures are not worthy of our sympathy. Although we might pity the immense ordeal they endure to please others, sympathizing with them seems to overly romanticize their servility while overlooking the fundamentally deplorable nature of their moral character.

According to Kant, "we must sympathize with the other's right, but not with the physical ill" (AL-F 25:606). This same principle can be applied to these individuals without glorifying their servility. Kant observes that "a fawning servility is likewise not indifferent to me" (LE 27:342). Considering that Kant also views misanthropy as a vice and to avoid the pitfall of victim-blaming, sympathy can be seen as the appropriate response to their servile acts. Such sympathy operates on two levels: first, I sympathize with the fact that they fail to recognize their own personal dignity as so precious that they would trade it as a commodity.

Second, when we witness servility in others, our sympathy does not seek to justify their actions; rather, it constitutes a sense of disappointment towards the thwarted moral expectation—they ought not to have acted this way. What we sympathize with is none other than the supreme human dignity that they have actively trampled upon, yet which ought to have remained inviolable. Far from contradicting our moral judgment, the capacity for sympathy, in this Kantian sense, in and of itself conveys our disapproval of the servile conduct.

[25] I agree with Hay (2011: 34) that while internal resistance can protect one's rational nature from oppression, it leaves oppressive structures intact; external resistance is therefore necessary.

Conclusion

I have been emphasizing a neglected aspect of Kant's duty of self-respect: that self-respect necessarily involves respect for others, and thus a concern for the humanity of all people is internal to self-respect itself. In light of human fragility, we sometimes feel *powerless* in defending both our own dignity and that of others, a powerlessness stemming partly from social injustice and partly from a lack of moral courage. In either case, even when we occasionally experience moral failure, we must not lose confidence and courage in our own moral capacity, for doing so would plunge us into what Kant describes as the cowardly form of servility. More importantly, self-respect is essentially the practice of regarding oneself as an equal among human beings. Because this equal status demands recognition from both oneself and others in the very act of interaction, self-respect necessarily entails respect for others.

References

Aquinas, T. 1920. *Summa Theologica*. Translated by Fathers of the English Dominican Province. New Advent. http://www.newadvent.org/summa/.

Bacin, S. 2013. "The Perfect Duty to Oneself Merely as a Moral Being (TL 6:428–437)." In A. Trampota, O. Sensen, and J. Timmermann (eds.). *Kant's "Tugendlehre"*. Berlin/Boston: De Gruyter, 245–268.

Baron, M. 1985. "Servility, Critical Deference and the Deferential Wife." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 48 (3): 393–400.

Boxill, B. R. 1995. "Self-Respect and Protest." In R. S. Dillon (ed.). *Dignity, Character and Self-Respect*. New York: Routledge, 93–104.

Boxill, B., and Boxill, J. 2015. "Servility and Self-Respect: An African-American and Feminist Critique." In M. Timmons and R. N. Johnson (eds.). *Reason, Value, and Respect: Kantian Themes from the Philosophy of Thomas E. Hill, Jr.* Oxford: Oxford University Press, 19–41.

Bunch, A. 2014. "Throwing Oneself Away: Kant on the Forfeiture of Respect." *Kantian Review* 19 (1): 71–91.

Blöser, C. 2015. "Degrees of Responsibility in Kant's Practical Philosophy." *Kantian Review* 20 (2): 183–209.

Cureton, A. 2013. "From self-respect to respect for others." *Pacific Philosophical Quarterly* 94 (2): 166–187.

Darwall, S. L. 1977. "Two Kinds of Respect." *Ethics* 1: 36–49.

Dean, R. 2006. *The Value of Humanity in Kant's Moral Theory*. Oxford: Oxford University Press.

Denis, L. 1997. "Kant's Ethics and Duties to Oneself." *Pacific Philosophical Quarterly* 78: 321–348.

———. 2015. "Proper Self-Esteem and Duties to Oneself." In L. Denis and O. Sensen (eds.). *Kant's Lectures on Ethics: A Critical Guide*. Cambridge: Cambridge University Press, 205–222.

Dillon, R. S. 2004. "Kant on Arrogance and Self-Respect." In C. Calhoun (ed.). *Setting the Moral Compass: Essays by Women Philosophers*. Oxford: Oxford University Press, 191–216.

Fahmy, M. S. 2013. "Understanding Kant's Duty of Respect as a Duty of Virtue." *Journal of Moral Philosophy* 10: 723–740.

Guyer, P. 2010. "Moral Feelings in the Metaphysics of Morals." In L. Denis (ed.). *Kant's Metaphysics of Morals: A Critical Guide.* Cambridge: Cambridge University Press, 130–151.

Hill, Jr., T. E. 1973. "Servility and Self-Respect." *The Monist* 57 (1): 87–104.

———. 2021. *Beyond Duty: Kantian Ideals of Respect, Beneficence, and Appreciation*. Oxford: Oxford University Press.

Hay, C. 2011. "The Obligation to Resist Oppression." *Journal of Social Philosophy* 42 (1): 21–45.

Holberg, E. A. 2017. "Kant, Oppression, and the Possibility of Nonculpable Failures to Respect Oneself." *The Southern Journal of Philosophy* 55: 285–305.

Johnson, R. N. 2010. "Duties to and regarding others." In L. Denis (ed.). *Kant's Metaphysics of Morals: A Critical Guide.* Cambridge: Cambridge University Press, 192–209.

Khader, S. J. 2021. "Self-Respect under Conditions of Oppression." In R. Dean and O. Sensen (eds.). *Respect: Philosophical Essays*. Oxford: Oxford University Press, 232–251.

Massey, S. J. 1983. "Kant on Self-Respect." *Journal of the History of Philosophy* 21: 57–73.

Møller, S. 2024. "What Is Humanity?" In L. Filieri and S. Møller (eds.). *Kant on Freedom and Human Nature*. New York: Routledge, 45–58.

Russell, F. 2020. "Kantian Self-Conceit and the Two Guises of Authority." *Canadian Journal of Philosophy* 50 (2): 268–283.

Stark, C. A. 2022. "Gaslighting, Self-Respect and the Kingdom of Ends." In J.-W. van der Rijt and A. Cureton (eds.). *Human Dignity and the Kingdom of Ends*. New York: Routledge, 266–282.

Timmermann, J. 2006. "Kant's Duties to the Self, Explained and Defended." *Philosophy* 81: 505–530.

Timmons, M. 2017. *Significance and System: Essays on Kant's Ethics*. Oxford: Oxford University Press.

——. 2021. *Kant's Doctrine of Virtue: A Guide.* Oxford: Oxford University Press.

Willaschek, M. 2024. "Kant's Answer to the Question 'What Is the Human Being'?" In L. Filieri and S. Møller (eds.). *Kant on Freedom and Human Nature*. New York: Routledge, 30–44.

Wood, A. W. 1999. *Kant's Ethical Thought*. Cambridge: Cambridge University Press.

——. 2008. *Kantian Ethics.* Cambridge: Cambridge University Press.