

Representation versus emancipation in modelling interactive human kinds

Caterina Marchionni and Marion Godman

15.03.2024

Abstract: An often-voiced concern about emancipatory approaches to modelling human kinds is that they are unlikely to reach their goals unless they rely on accurate knowledge of the kinds they target; knowledge, it is assumed, which can only be obtained by representing the kinds as accurately as possible independently of any particular social or political goal. We argue that this argument is problematic for several reasons. We show that even if the pursuit of emancipation should indeed rely on accurate knowledge about kinds, a merely representational approach is neither necessary nor sufficient to obtain such knowledge.

1. Introduction

How should we model human kinds like gender and race? Should social scientists rather be committed to promoting, through their modelling efforts, political and social values grounded on ideals of social justice and emancipation? Or should they aim at representing them - the mechanisms that bring them about, the properties they have, and the generalizations in which they partake – as accurately as possible?

The philosophical debate on human kinds has traditionally focused on their ontology, such as the putative demarcation from, or reduction to, natural kinds, and their resulting epistemology (e.g. Hacking 1986, 2007; Boyd 1991; Millikan 1999; Mallon 2016; Godman 2020). In this paper we draw on what we take to be some lessons of these debates: first, even if human kinds are socially constructed in an obvious, it is nevertheless possible to have knowledge about them, the mechanisms that bring them about and sustain them, and the generalizations in which they partake; second, human kinds are interactive, i.e.- the kinds undergo change in response to interactions between people (and their awareness) and categorisation (see e.g. Hacking 2007; Khalidi 2010) and relatedly, they are infused with value – they are kinds that interact with what we are and want to be like (e.g. Hacking 1995).

Our question is different from the standard ones in these debates, however. We ask, given that some human kinds are interactive and often highly connected to value, identity and injustice how should they be studied? We distinguish two ways of modelling human kinds: one, which we call *emancipatory*, that aims at modelling human kinds in ways that promote emancipatory goals; the other, more traditional approach, which we call *representational*, aims at describing human kinds, their characteristics, and the generalisations in which they partake as accurately as possible independently of any particular political goal.

Inspired by Bach (2019)'s terminology, we talk of them as models. We use the term *models* broadly to cover any systematic inquiry into human kinds, not only those that involve the type of surrogate reasoning thought of as distinctive of scientific modelling as an epistemic activity (e.g., Frigg and Hartman 2020). Calling them models should be a reminder that such inquiries typically involve aspects of both representation and construction, since modelling always involve making choices about which features of a target to include and which to exclude.

We then turn to an argument often advanced in favour of representational approaches and against emancipatory ones, namely that the latter are bound to fail in their emancipatory goals (i) unless they rely on knowledge of the kinds, the mechanisms that bring them about and sustain them, and the generalizations in which they partake; (ii) knowledge, in turn, which can only be obtained by committing to a fully representational approach to. We refer to this argument as the *instrumental reliability argument* since it is a concern about what is the most reliable means to promote social and political goals like emancipation. In other words, it is a concern about whether or not such goals can be fulfilled by letting them guide inquiry from the get go as it's the case for emancipatory modelling.

We show that, in spite of its appeal, the instrumental reliability argument is flawed in three respects. First, emancipatory approaches can produce reliable knowledge that can be used for purposes of prediction and intervention in spite of (or even thanks to) their explicit commitment to social and political values. Second, because human kinds are interactive, a representational approach raises the same kind of concerns about instrumental reliability as the emancipatory approach. For human kinds, any scientific study has the potential to affect their properties, the mechanisms that maintain them, and the generalizations in which they partake. Hence, the causal relations behind and around such kinds are fragile in ways that make both prediction and intervention difficult in general. Third, for reactive human kinds, modellers ought to take cognizance of the potentially harmful effects of their models in ways that require social and moral considerations to guide, or at least constrain, epistemic ones.

Our focus, in what follows, will mainly be on modelling gender, but our arguments generalize to any other human kinds that people care about and are the subject of modelling in science, philosophy, and related areas such as diagnostics, journalistic and policy enquiries. Moreover, gender has been a focus of the different bodies of literature that we bring together in this paper: the literature on emancipatory and ameliorative modelling (e.g., Haslanger 2012; Jenkins 2016), the most recent literature on interactive human kinds (e.g., Laimann 2020; Peters 2023), and finally the burgeoning literature on values and science (Longino 1990, Douglas 2000).

2. Emancipatory approaches to modelling human kinds

A common trend across different bodies of scholarship, including post-colonial studies, social and critical theory, disability, gender, and indigenous studies, is to critically examine and model the human kind in question in terms of achieving greater emancipation from certain institutional structures and other processes deemed unjust.¹

In the words of critical social theorist Erik Colin Wright (2010, 11):

It is not enough to show that people suffer in the world in which we live or that there are enormous inequalities in the extent to which people live flourishing lives. A scientific emancipatory theory must show that the explanation for this suffering and inequality lies in specific properties of institutions and social structures. The first task of emancipatory social science, therefore, is the diagnosis and critique of the causal processes that generate these harms.

¹ This can be an emancipation from one or several of the following injustices: historical injustices (e.g. past oppressions and colonialism), structural injustices (e.g. institutional racism, capitalism and other discriminatory practices), and epistemic injustices (e.g. silencing and obstruction of testimony).

In the field of disability studies, we find a similar programmatic statement from Colin Barnes (1992: 122):

Emancipatory research is about the demystification of the structures and processes which create disability, and the establishment of a workable dialogue between the research community and disabled people.

The aim of emancipatory research is not primarily that of describing social reality as accurately as possible. The goal is to describe social reality in ways that are most likely to contribute to countering or rectifying real-world injustices and lead to greater emancipation.

For example, in Stewart's et al.'s study of divorce (1997) the commitment to feminist values of equality and emancipation shows up in separating spousal and parental roles attached to gender. Divorce is seen as an opportunity for personal growth rather than merely as a loss or trauma (Stewart et al. 1997, 19). As such, it contributes to promote the emancipation of women and children (and possibly also men) by means of a modelling effort that both analyses and departs from traditional family structures where parenting and spousal roles are closely linked.

A commitment to the promotion of social justice and emancipation is also a core feature of Sally Haslanger (2000, 2012)'s influential *ameliorative analyses of gender and race*. Haslanger recommends that moral and political concerns should guide the modelling of gender and race: "A primary concern of feminist and antiracist theorizing is to give an account of the social world that will assist us in the struggle for justice (2012, p. 6)." In the service of fighting gender oppression specifically, Haslanger proposes that gender should be defined according to one's position in a hierarchical society where certain groups are privileged, and others are not. She suggests it is useful to attend to and question how people's social identities are formed within a given hierarchy and proposes a classification in explicit recognition of such hierarchical identities: women should be defined according to their relative subordination and men according to their relative privilege or domination along different societal dimensions.²

3. The instrumental reliability objection

Several worries have been raised again emancipatory approaches across both philosophy and social science (for early critiques, see e.g. Haack 1993; Gross and Levitt 1994). The worry we focus on states that even if we accept that the political goals of emancipatory approaches are worthy of pursuit, putting them first in empirical inquiry is not the best way to achieve those aims. On the contrary. There is no guarantee that emancipatory approaches will succeed in achieving their aims, and worse, they may negatively affect the very people they're trying to emancipate (Bach 2019, 2022; Lacey 2002; Saul 2006).

For example, Jennifer Saul (2006, 138) points out that Haslanger's proposed definitions of gender might backfire:

Those in subordinated positions might instead become trapped in a feeling of powerlessness to change their own fates. It could be quite disempowering, for example,

² Haslanger similarly attempts to give a socio-political account of races as groups who are either observed or imagined as having certain bodily features related to geographical ancestry and that are used as markers for practices of subordination and privilege (2012, pp. 308).

for a woman to come to believe that women are by definition subordinated. And those in powerful positions might feel even more secure and deserving of their power if they come to see it as built into being (for example) a man. Certainly such sentiments are not at all unheard-of amongst the subordinated and the subordinators. Whether these responses or those Haslanger suggests would occur is a matter of human psychology. Gambling on the positive responses Haslanger expects is risky.

The idea here is that modelling gender kinds in the fashion suggested by Haslanger could interact with human psychology in ways that might counteract the intended effects, to the point of harming rather than helping the people the models were supposed to assist in the first place (see also Mikkola 2016: 84-86). This would clearly be not only an unintended but an undesired effect of Haslanger's type of ameliorative analysis. The possibility of counterproductive effects may also be present in the case of modelling divorce. For example, we might imagine that emphasising the opportunity for personal growth when modelling divorce once disseminated may lead politicians and policymakers to downplay the institutional and social support divorced people may actually need. Hence the commitment to emancipation ends up worsening rather than improving the situations of many women.

By itself this only shows that the world-changing aims of emancipatory models might be self-defeating, not that the whole approach is misguided. To make that claim, another premise is needed, namely, there is another way of studying human kinds that gives us better chances of achieving the aims that emancipatory modellers set for themselves. Theodor Bach (2019, 2022) pursues this line of argument. He argues that the *responsible modelling* of human kinds involves tracking the kinds and in particular the causal mechanisms behind the clustering of properties of the kind. Accurate representation is, according to Bach, a precondition for successful interventions such as changing the properties of existing human kinds (e.g., allowing women to be better at spatial rotation tasks), and even getting rid of the kinds (i.e., eliminating gender as a socially relevant kind). Intervention however is in the purview of political projects, not scientific ones. According to Bach, the possibility that emancipatory models may backfire demonstrates that that the political goal of emancipation will only be attainable if researchers commit to what we call a representational approach. In other words, only a representational approach to the study of human kinds can give us knowledge that can be used to predict and intervene to promote political goals such as emancipation.

4. Emancipatory approaches and objective knowledge about kinds

The instrumental reliability argument is quite convincing: we do need accurate or we may say objective knowledge about how the world works to effectively intervene in it.³ On the other hand, Bach (2019) and other supporters of the instrumental reliability argument seem to assume that emancipatory modellers cannot avail themselves of the knowledge about kinds that would be needed to improve the world-changing prospects of their models – that is, accurate knowledge about clotting mechanisms, properties, and generalisations that can be used for purposes of prediction and intervention. But why not?

Being committed to emancipatory values does not necessarily entail wishful thinking, dogmatism or relativism (Alcoff 1987). Recent literature on values and science has convincingly showed that it is possible to let political and social values guide one's scientific

³ We don't distinguish between accuracy, reliability, and objectivity in this paper: all three concepts roughly point to the idea that we need to get as close as possible to the real mechanisms behind kinds in order to be able to use knowledge about them for purposes of prediction and intervention.

projects while being sensitive to evidence as well as ready to revise one's own value commitments if our empirical inquiries led us to do so (e.g. Anderson 1995, Longino 1990)). These arguments are well-known. Briefly: any scientific inquiry involves making choices about what to include and what to leave out, and how to represent what is included. That these choices are partly motivated by political values does not detract from the possibility that the knowledge we obtain about that slice of the social world is objective. Elizabeth Anderson (1995) famously argues that science does not pursue just any truth, but *significant* truths, and judgments of significance are partly judgments of values, which are themselves sensitive to empirical matters (see also Kourany 2016, Brown 2022). The commitment to emancipation guides scientists to carve the social world in ways they judge more likely to promote emancipation, not as a substitute to evidence. Accordingly, there may be equally legitimate ways of drawing the boundaries of human kinds, single out the relevant clusters of properties and hence the mechanisms that bring about and maintain them.

This makes sense of why emancipatory researchers can comfortably claim allegiance to accurate description while being explicitly guided by political values. Stewart et al's study is indeed grounded in robust empirical evidence. In reviewing it, Anderson (2004) suggests that values grounded in emotional experiences directed the researchers toward a different type of qualitative data that ultimately proved useful for improving perceptions of divorce. Haslanger has been explicit about the need to tie emancipatory models to social reality (e.g. 2015). In fact, her ameliorative project can be interpreted as aiming to capture the causal mechanisms behind gender for example. There definitely are clotting mechanisms that would account for why members of the kind *woman*, understood as the subordinate gender, tend to display the same characteristics – even if on this account, the relevant cluster is different from other ways of conceptualizing and studying the kind. Such clotting mechanisms may also explain why self-identified women do share many of the properties of the kind woman as Bach and others understand it.

Furthermore, when our emancipatory modelling looks unlikely to have the desired effect or when its value assumptions stand in need of revision – the cases that have quite rightly worried many for their possibly counterproductive effects – the emancipatory goal should motivate revisions of one's modelling efforts. After all, if there is truly a commitment grounded in the ideal of emancipation, the modellers should be ready to adjust and revise both modelling assumptions and value commitments in light of empirical evidence.

Thus, emancipatory modelling need not mean a commitment to a specific model or assumption that is somehow insulated from the predicted effects on the kind or on the world. One the contrary. If Haslanger's approach to gender and race turned out to be unlikely to be ameliorative, then the model should supposedly be revised. The same goes more broadly for conceptual engineering and other emancipatory approaches: if the engineering or modelling efforts look likely to be self-defeating, it is precisely the goal or purpose of the engineering to guide their revision. The ethos of emancipatory approaches to modelling human kinds can, or rather, should have precisely a reason to be accurate and revise its approach in light of empirical evidence.

5. The challenge of interactive human kinds

Arguing that an emancipatory approach to modelling human kinds is compatible with having accurate knowledge of the kinds does not fully address the instrumental reliability concern. After all, it is still possible that emancipatory models, in their well-meant attempts to promote emancipation, will backfire since the way members of the kind will respond to being

modelled in a certain way is often hard to predict. The worry here is epistemic: it's about the possibility of predicting the effects that a certain way of modelling a kind has on people of that kind. This is not an easy feat. It involves predicting whether and how the model will be picked up by the relevant agents, how it will be interpreted, and how people of the kind as well as the people and institutions around them will react.

Hence, the worry is real. Our claim is not that emancipatory modelling does not run the risk of backfiring, but rather than it is a risk that any approach to modelling human kinds ought to face – whether or not their primary intent is to change the reality they model. The general problem stems from the fact that many human kinds are interactive (Khalidi 2010) individuals falling under a scientific human categorisation can change in response to being studied and modelled, thereby also changing the kind and its properties, making them *moving targets* (Hacking 1995). Accordingly, what is known about a kind at a given point in time may be of limited use for the purposes of predicting and intervening into a kind whose features may change as a result.

One reason why Hacking highlights the interactive effects of scientific classifications is the institutional power and epistemic authority vested in science – often coupled with equally influential bureaucratic practices surrounding diagnostics and education (Hacking 1995; 2007). For example, many scientific claims about gender are incorporated into policy recommendations and directives concerning whether single sex schooling or, in contrast, of gender-neutral preschooling, causing interactive effects in these venues. Other scientific claims and testimony about gender and gender differences tend to be mediated by and picked up from popular science, traditional media, social media, and self-help manuals which sometimes amplify, sometimes simplify, and sometimes skew scientific claims. In fact, the most powerful scientific claims about gender are arguably those which combine the features of being central to scientific investigation, being widely disseminated, and regarded as more settled by the general public than others.

Consider the generalisation that “women are more risk averse than men” as well as the explanation typically associated with it, namely, the stable or natural differences in testosterone levels between women and men. Innate or natural hormonal differences in testosterone levels are thought to explain typical male traits like increased aggression, violence and general strategies that involve risk taking, while low testosterone is thought to be correlated with typical female traits like caregiving (see e.g. Cueva et al 2015 and Herbert 2015). This kind of claims have been used to explain why men tend to dominate in competitive situations such as sport, as well as in professional settings where risk-taking is encouraged, such as finance, banking, and politics (Nierdele & Vesterlund 2011; Preece et al 2015).

It is relatively established that scientific claims (and their popularisation) may influence individuals into conforming behaviour. For example, high levels of anxiety lead some individuals to adopt less risky behaviours in accordance with (female) gender stereotypes (Schmader 2010). That there are such *conformity effects* is a common theme in work on stereotype threat (Dar-Nimrod & Heine 2006; Schmader 2010). Because of conformity effects a scientific generalisation such as “women are more risk averse than men” might not be the result of hormonal differences but of conformity effects.

Another way in which generalisations about gender differences could be stabilised is via the introduction and use of *biological or nativist* explanations. There is convergent evidence from

many different parts of psychology that citing genetic or brain-based explanations is associated with the idea that the kind or trait is fixed, irreversible or outside our control. When psychiatric symptoms are attributed to the brain they are considered less within the patient's control (e.g., Deacon & Baird, 2009; Kemp, Lickel, & Deacon, 2014). This research points toward the mere biologizing of traits leading to *stabilising effects* on the generalisation, as any occurrence of a supposed biologically-based trait tends to be conjoined with the thought that it was inevitable and that it cannot be changed.

But interactive effects may also run contrary to this pattern such as when people resist or reject certain scientific claims about them. Sarah Townsend and colleagues (2011: 151) for instance showed that there are “powerful interactive effects” of chronic perceptions of sexism. Indeed, these authors tracked measurable neuroendocrine levels of stress and found that it was raised among women attentive to sexism compared to women who were not. Certainly, the critical attention to claims about women's innate risk aversion indicate that this might be the kind of situational cue that sets off resistance that destabilises the generalizations.

Not only human kinds undergo changes in response to being investigated, but such changes occur in a wayward manner and are often quite hard to predict beforehand (Laimann 2020). This is because the causal relations behind and about human kinds tend to be fragile while we currently lack a theory of human psychology that can tell us under which conditions they are robust to contextual changes (Northcott 2022). When it comes to interactive human kinds representational models therefore do not necessarily place us in a better position for purposes of prediction and control. As for emancipatory models, representational models too may have unintended and harmful effects on people, creating another layer of complexity; one where social and political values come in from the backdoor.

6. Dealing responsibly with unintended harmful consequences

We have seen that models of human kinds can change the kinds themselves, and/or the mechanisms which maintain them, and the generalisations in which they partake. Some changes endanger the reliability of the knowledge we have about them. Scientists will have an interest in tracking these changes for epistemic reasons. Some of these changes may also be harmful. For example, scientific claims about women cognitive abilities or low testosterone levels may reinforce existing prejudices and reduce the space of agency and freedom of those the claim is about (e.g. Kourany 2016). In such cases, there are also non-epistemic reasons to worry about the world-changing features of models of human kinds too.

Insofar as scientists have responsibilities towards harmful effects of their scientific claims and models in virtue of their special power and epistemic authority, concern for those unintended harmful effects is shared across both representational and emancipatory modellers (Douglas 2003, 2009, Resnik 1998, Carrier 2021). As for inductive risk more generally, dealing responsibly with the possibility of harmful consequences from the modelling of interactive human kinds requires the assistance of non-epistemic values. Heather Douglas (2000, 2003) explains that such decisions cannot be entirely handed over to outside actors as this would amount to scientists renouncing their hard-won autonomy. At any stage of inquiry one can make errors whose consequences may have impacts well beyond the scientific pursuit, making the possibility of relinquishing value judgments not only undesirable but also not practicable. Given the possibility of interactivity, this clearly applies

also to representational modelling. Thus, any modelling of interactive human kinds is likely to require the guidance of non-epistemic values.

Before concluding, we should address a remaining objection. It could be argued that even if both kinds of modelling approaches can trigger interactive effects (and hence that non-epistemic values may need to be consulted at different stages of the inquiry), it is only thanks to accurate representation that we will be in a position to determine when and how a model produces such effects and what to do about it. This, the objection goes, shows once again that representational modelling is better placed to give us the handles with which we can promote political goals, including emancipation. In other words, to address interactivity, for both epistemic and non-epistemic purposes, investigation of the interactive mechanisms between science and human kinds is needed to tell when interactivity occurs, its direction and how to deal with it.

Our reply is that even when this kind of information can be had (and given the wayward manner in which human kinds often behave this is not always the case), why should information not be available to the emancipatory modeller? To the extent that systematic investigation of the world-changing features of scientific claims is possible, that should be available to emancipatory approaches as well. There is no reason to suppose that emancipatory modellers should carry out these investigations any differently than representational modellers. For some inquiries, taking an emancipatory or a representational approach might in practice amount to making the same choices. Does this mean that after all there aren't two equally legitimate ways of modelling interactive human kinds, but only one that aims at accurate representation while keeping the role of political goals such as emancipation only *contextual* as opposed to *constitutive* (Douglas 2000)? At this level of abstraction probably yes. But being explicitly committed to the promotion of the political goals of emancipation from the start (versus some other political goal or to no political goal at all) may still make a relevant difference at the level of methodological choices. Now that the instrumental reliability argument is out of the way, we can more productively discuss what these different choices are or should be.

7. Concluding remarks

We have argued that the instrumental reliability argument in favour of a representational approach and against an emancipatory approach to modelling human kinds is flawed. The argument is especially problematic when we are concerned with human kinds that are potentially interactive. Debunking the instrumental argument does not amount to a full-fledged defence of an emancipatory approach. Among other things, that would require an independent argument in favour of the promotion of emancipation vis-à-vis other political values or moral principles (e.g. doing no harm). Is emancipation really the overarching ideal to pursue when modelling human kinds? Setting aside the one-sided concern about emancipatory models being likely to be self-defeating to one side allows us to address normative concerns about how to best balance epistemic and epistemic goals in the modelling of interactive human kinds.

References

- Alcoff, L. (1987). Justifying feminist social science. *Hypatia*, 2(3), 107-120.
- Anderson, E. (1995). Knowledge, human interests, and objectivity in feminist epistemology. *Philosophical topics*, 23(2), 27-58.
- Bach, T. (2019). Real Kinds in Real Time: On Responsible Social Modeling. *The Monist*, 102(2), 236-258.
- Bach, T. (2022). Same-tracking real kinds in the social sciences. *Synthese*, 200(2), 118.
- Barnes, C. (1992). Qualitative research: valuable or irrelevant? *Disability, handicap & society*, 7(2), 115-124.
- Boyd, Richard. "Realism, anti-foundationalism and the enthusiasm for natural kinds." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 61.1/2 (1991): 127-148.
- Brown, M. J. (2020). *Science and moral imagination: A new ideal for values in science*. University of Pittsburgh Press.
- Carrier, M. (2019). How to conceive of science for the benefit of society: prospects of responsible research and innovation. *Synthese*, 1-20.
- Cueva, Carlos, R. Edward Roberts, Tom Spencer, Nisha Rani, Michelle Tempest, Philippe N. Tobler, Joe Herbert, and Aldo Rustichini. "Cortisol and testosterone increase financial risk taking and may destabilize markets." *Scientific reports* 5, no. 1 (2015): 1-16.
- Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science*, 314(5798), 435-435.
- Deacon, B. J., & Baird, G. L. (2009). The chemical imbalance explanation of depression: Reducing blame at what cost? *Journal of Social and Clinical Psychology*, 28(4), 415-435.
- Douglas, H. (2000). Inductive risk and values in science. *Philosophy of science*, 67(4), 559-579.
- Douglas, H. E. (2003). The moral responsibilities of scientists (tensions between autonomy and responsibility). *American Philosophical Quarterly*, 40(1), 59-68.
- Douglas, H. (2009). *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press.
- Fine, C. (2017). *Testosterone Rex: Unmaking the myths of our gendered minds*. Icon Books.
- Frigg, Roman and Stephan Hartmann, "Models in Science", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2020/entries/models-science/>>.
- Godman, M. (2020). *The Epistemology and Morality of human kinds*. Routledge.
- Gross, P. R., & Levitt, N. (1997). *Higher superstition: The academic left and its quarrels with science*. JHU Press.
- Haack, S. (1993). Epistemological reflections of an old feminist. *Reason Papers*, 18.

Hacking, I. (1986), 'Making Up People', in Heller, Sosna, and Wellbery (eds), *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought*, Stanford: Stanford University Press, pp. 222-236.

Hacking, Ian. 1995a. "The looping effects of human kinds". In *Causal cognition: A multidisciplinary debate*, Edited by: Sperber, Dan, Premack, David and Premack, Ann J. Oxford: Clarendon Press.

Hacking, I. (2007). Kinds of people: Moving targets. In *Proceedings-British Academy* (Vol. 151, p. 285). Oxford University Press.

Haslanger, S. (2000). Gender and race:(What) are they?(What) do we want them to be? *Noûs*, 34(1), 31-55.

Haslanger, S. (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.

Haslanger, S. (2015). Theorizing with a purpose: The many kinds of sex. In *Natural kinds and classification in scientific practice* (pp. 129-144). Routledge.

Heine, S. J., Dar-Nimrod, I., Cheung, B. Y., & Proulx, T. (2017). Essentially biased: Why people are fatalistic about genes. In *Advances in experimental social psychology* (Vol. 55, pp. 137-192). Academic Press.

Herbert, J. (2015). *Testosterone: Sex, power, and the will to win*. OUP Oxford.

Jenkins, K. (2016). Amelioration and inclusion: Gender identity and the concept of woman. *Ethics*, 126(2), 394-421.

Khalidi, M. A. (2010). Interactive kinds. *The British Journal for the Philosophy of Science*.

Kemp, J. J., Lickel, J. J., & Deacon, B. J. (2014). Effects of a chemical imbalance causal explanation on individuals' perceptions of their depressive symptoms. *Behaviour Research And Therapy*, 56, 47-52. doi:10.1016/j.brat.2014.02.009.

Kourany, J. A. (2016). Should some knowledge be forbidden? The case of cognitive differences research. *Philosophy of Science*, 83(5), 779-790.

Kourany, J. A. (2020). The New Worries about Science. *Canadian Journal of Philosophy*, 1-19.

Lacey, Hugh. "Explanatory critique and emancipatory movements." *Journal of Critical Realism* 1.1 (2002): 7-31.

Laimann, J. (2020). Capricious kinds. *The British Journal for the Philosophy of Science*, 71(3), 1043-1068.

Mallon, R. (2016) *The Construction of Human Kinds*. Oxford University Press.

Millikan, R. G. (1999). Historical kinds and the "special sciences". *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 95(1/2), 45-65.

Mikkola, M. (2016). *The Wrong of Injustice: Dehumanization and Its Role in Feminist Philosophy*, Oxford: Oxford University Press.

Northcott, R. (2022). Reflexivity and fragility. *European Journal for Philosophy of Science*, 12(3), 43.

- Peters, U. (2023). Science communication and the problematic impact of descriptive norms. *The British Journal for the Philosophy of Science*, 74(3), 713-738.
- Pinder, M. (2022). Is Haslanger's ameliorative project a successful conceptual engineering project?. *Synthese*, 200(4), 334.
- Preece, J., & Stoddard, O. (2015). Why women don't run: Experimental evidence on gender differences in political competition aversion. *Journal of Economic Behavior & Organization*, 117, 296-308.
- Saul, J. (2006) Gender and Race. In *Aristotelian Society Supplementary Volume* (Vol. 80, No. 1, pp. 119-143). Oxford, UK: Oxford University Press.
- Scheff, Thomas (1966). *Being Mentally Ill: A Sociological Theory*. Aldine, Chicago.
- Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science*, 19(1), 14-18.
- Townsend, S. S., Major, B., Gangi, C. E., & Mendes, W. B. (2011). From "in the air" to "under the skin": Cortisol responses to social identity threat. *Personality and Social Psychology Bulletin*, 37(2), 151-164.
- Wright, E. O. (2010). *Envisioning real utopias* (Vol. 98). London: Verso.