

Abstract: The social categories to which we belong—*Latino, disabled, American, woman*— causally influence our lives in deep and unavoidable ways. One might be pulled over by police because one is Latino, or one might receive a COVID vaccine sooner because one is American. Membership in these social categories most often falls outside of our control. This paper argues that membership in social categories constitutes a restriction on human agency, creating a situation of non-ideal agency for many human individuals.

However, there are ways to resist the causal influence of social categories, and certain socially marginalized groups can be understood as attempting to do just this. I discuss two instances of social category resistance: gender pronouns and the rights of trans individuals. I suggest that the intentional declaration of gender pronouns (“she/her” or “they/them”) can be understood as an attempt to resist the causal powers of social categorization. Similarly, one among many reasons to support the rights of trans individuals is that their self-declaration of gender identity can be viewed as a reclamation of agency in the face of causal constraints imposed by socially defined and imposed gender categories. This lesson can be generalized to people belonging to a broad range of marginalized groups.

Keywords: social categories, social metaphysics, causation, oppression, philosophy of gender, philosophy of race, agency

Resisting Social Categories

Sara Bernstein, University of Notre Dame

(forthcoming in *Oxford Studies in Agency and Responsibility* Volume 8, eds. Shoemaker, Amaya, and Vargas—please cite published version)

The social categories to which we belong—*Latino, disabled, American, woman*— influence our lives in deep and unavoidable ways. One might be pulled over by police because one is Latino, or one might receive a COVID vaccine sooner because one is an American citizen. Membership in these social categories most often falls outside of our control: though we can choose “how we identify,” we can’t opt out of the causal influence of externally-imposed social categories on our practical and moral lives.

This paper argues that membership in social categories constitutes a restriction on human agency (broadly construed), creating a situation of non-ideal agency for most human individuals. However, there are ways to resist the causal influence of social categories, and certain socially marginalized groups can be understood as attempting to do just that.

Roadmap: In Section 1, I begin with a brief overview of the nature of social categories. I draw on recent work in analytic metaphysics to support the claim that social categories are the sorts of things that can be causal. Whereas it is sometimes thought that only the most ontologically basic things are causal, I suggest that social categories are the sorts of things that can cause other things to happen. This is, in part, because social categories can be thought of as ontologically fundamental, on a certain conception of fundamentality. I provide several compelling examples of social categories as fundamental causal *relata*. Social categories cause things, in part, by furnishing social reasons which mediate social reality.

In Section 2, I argue that the causal power of social categories constitutes a special sort of restriction on human agency. Since membership in most social categories is not something that can be chosen, and since the natures of social categories are extremely difficult to modify, many social categories infuse our lives in ways we cannot choose or change. While people do sometimes shift their membership in social categories, the most causally important social categories to which people belong often remain unchanged throughout their lives. As membership in social categories is often a matter of luck, a large swath of life experience is caused by lucky or unlucky social categorization outside of one's control. This sort of taxonomic luck is a form of Nagelian constitutive moral luck.

In section 3, I discuss what it is to resist social categories given the constraints they place on our agency. I lay out several different forms of social category resistance: changing the social categories to which one belongs, changing the causal profile of social categories, and changing the conceptual content of social categories. I suggest that a recent trend in social philosophy, conceptual engineering, is a form of social category resistance. While these forms of resistance are ways to reclaim agency, all forms of resistance face numerous obstacles to being vehicles of social change.

In section 4, I apply the results to three timely topics in the philosophy of gender: gender pronouns, trans rights, and transracialism. I suggest that the intentional declarations of gender pronouns (“she/her” or “they/them”) can be understood as attempts to resist external social categorization and its causal power. Similarly, one among many reasons to support the rights of trans and nonbinary individuals is that their self-declarations of gender identity can be understood as a reclamation of agency in the

face of causal constraints imposed by gender categories. This lesson can be generalized to people belonging to a broad range of marginalized groups.¹ On the flipside, there are some cases of social category resistance that are morally problematic, including cases of transracialism.

1. Social Categories are Causal

“Social category” is a catch-all term for a complex phenomenon. Distinguish between external and internal social categories. External social categories are those to which one is generally perceived by others to belong. Roughly, if one is treated by one’s social community as a member of social category x in context c , one is a member of social category x in context c . What one is perceived to be varies by situation and social community; I won’t tend to these details here. For our purposes, we can assume that everyone is a member of external social categories, and that perception of membership in these categories is stable across a range of social situations.

In contrast, internal social categories are one’s “felt” social categories, or those to which one feels one belongs. Felt social categories are complex and heterogenous.² Conscious adoption of internal social categories is a highly variable matter throughout one’s lifetime: one might feel like a woman at one life stage and a man at another life stage. Or one might not fully accept oneself as Southeast Asian until late in life, even if one is consistently perceived to belong that category. One need not “feel” an internal social category at every moment in order to count as belonging to it: I don’t feel very Jewish in New York City, for example, but I feel extremely Jewish at the University of Notre Dame.

The line between external and internal social categories is messier than one might expect. For example, one might unreflectively consider oneself a woman just because everyone else perceives them to be a woman. Or a stereotype of Black masculinity might influence the felt social category of a Black man who does not satisfy the stereotype.

¹ See Webster (forthcoming) for a discussion of these broader issues.

² Dembroff and Saint-Croix outline a key sort of felt social category, agential social category, which involves “self identification and role-directed externality.” (2019, pp. 581) They define agential identities as “a relation between individuals, their self-identities, and social positions. They act as a bridge between one’s internal identification and one’s preferred public perception.” (ibid, p. 574)

While my arguments are broadly aimed at external social categories, I will not depend on a clean metaphysical distinction between external and internal social categories. Both sorts of categories dwell in the social world.

Social categories fall on a spectrum of granularity from coarse-grained (“women”) to fine-grained (“Black cisgender women born before 1968”). Social categories can also be highly localized and extremely fine-grained (“sophomore goth teenagers at Adams High School in South Bend, Indiana”). While all such social categories influence the lives of their members, I will be largely interested in coarse-grained social categories that carve macro-level social reality at its joints at a time: “Black women” is such an example; “sophomore goth girls at Adams High School in South Bend, Indiana” is not. For my purposes, the categories that carve social nature at its joints are those that back predictions and explanations in a particular cultural context.³

Social categories are inescapable-- we all belong to them.⁴ I am an American cisgender woman: my life is influenced in many ways by each of these categories, individually and collectively. Altering my membership in any of these categories would have altered my life trajectory. Casual reflection reveals a similar story for each of us: our lives are shaped by the categories to which we belong. Membership in social categories exerts extraordinary influence over each human life, in ways large and small.⁵

In my view, this is because social categories are causal: they function as causes, effects, and causal intermediaries. Here I will give the general thrust of an argument that I elaborate in detail elsewhere.⁶ We routinely think things like “If I had been born a man, I wouldn’t have faced early pressure to get married” and “If I hadn’t been a young woman, I would have an easier time in that negotiation.” Social categories centrally feature in causal claims like “Being femme causes Janelle to be mistaken for being straight and cis” and “DeShawn was pulled over because he was Black.” Such claims generate corresponding true causal counterfactuals like “If Janelle hadn’t been femme, she wouldn’t have been mistaken for being straight and cis” and “If DeShawn hadn’t been

³ Social categories change natures across times. Thanks to David Shoemaker for this point.

⁴ It is possible that there are completely isolated humans without any causally significant social category membership, like a person alone on a desert island. Here I set aside these cases.

⁵ For extensive discussion of this point in Black political thought, see Du Bois (1903), Hill Collins (1990), and Fanon (1952).

⁶ See my “Fundamental Social Causes” (manuscript).

Black, he wouldn't have been pulled over.” In cases such as these, counterfactually altering the social categories changes the outcomes.⁷ If Janelle had been transmasc, she wouldn't have been mistaken for being straight and cis; and if DeShawn had been white, he would not have been pulled over.

Social categories have *causal profiles*. Roughly speaking, causal profiles are the ranges of ways that the categories behave in the lives of their bearers.⁸ Causal profiles encompass what Ásta (2018) calls “constraints and enablements”: social categories can restrict what their members do, or enable what their members do.⁹ For example, the social category *high socioeconomic status white straight cis man* has a range of causal powers in the lives of its members: such persons are advantaged in various ways because of membership in this category, and (to a lesser extent) disadvantaged in others.¹⁰ Minoritized and marginalized social categories have constraining causal profiles: being Black in America, for example, produces social limits and restrictions on its members.¹¹ Causal profiles of social categories also intersect and interact: intersectional social categories like *Black woman* have different causal profiles than both categories added together.¹² As Ásta notes, the causal profiles of social categories are situationally variable. The causal constraints imposed by the category “woman” are different at a workplace and in a gay bar.¹³

Social reasons are a medium for the causal power of social categories.¹⁴ Reasons determine how people treat others, and they determine how one is treated. Social

⁷ These interventions have similar explanatory power for macro-level social causal claims.

⁸ Katharine Jenkins' concept of “ontic injustice” is a type of causal profile, on my view. See Jenkins (2020) for more discussion.

⁹ Thanks also to Manuel Vargas and Aness Webster for pointing out that social categories can assist their members as well as constrain them, depending on the context.

¹⁰ Here I am choosing not to cash this out in terms of dispositions, but see Bird (1998) and McKittrick (2004) for some relevant literature.

¹¹ As an astute referee notes, such categories do also enable certain forms of agency: for example, minoritized social groups often form supportive communities based on group membership. I will focus primarily on constraints on agency in this paper, though enablements of minoritized social categories are very important for understanding the natures of those categories, and the experiences of people in them.

¹² See my (2020) for an argument that intersectional social categories are explanatorily unified. See my (ms) for an argument that intersectional social categories have finer-grained causal detail than unitary social categories.

¹³ See Ásta (2018) pp. 15-18 for relevant discussion.

¹⁴ Here I intend social reasons to encompass both justificatory and explanatory reasons. Both sorts operate in the ways I suggest.

categories furnish social reasons to which people respond. For example, the social category *Black man in the United States* furnishes social reasons that mediate social and legal treatment of people who fall in this category. Without the social category, those particular social reasons wouldn't exist. Counterfactually altering the nature of a social category alters the social reasons furnished by it. For example, if Black men in the United States had occupied a non-marginalized social position, the reasons furnished by the social category *Black man* would be different than they are.

If human agency involves responsiveness to reasons, social agency involves responsiveness to social reasons.¹⁵ Reconsider the example of DeShawn being pulled over because he is Black. In this case, DeShawn's membership in a social category provides a reason that the police officer pulls him over. Similarly, Meena's marriage prospects as a Dalit woman are caused by people responding to social reasons. Aside from influencing our everyday treatment, social reasons also cause people to behave in certain ways: Meena chooses possible marriage partners based on her caste, for example, and DeShawn must be extra careful to go under the speed limit because he is Black. The social reasons that influence our lives—both those that we have and those that cause others to respond in certain ways to us—are furnished by social categories.¹⁶

It is tempting to resist the idea that social categories are causal. Because social categories are mind-dependent social constructions, they are canonically taken to be non-natural and non-causal. If humans didn't exist, then social categories wouldn't exist. So, the thinking goes, they are not the sorts of things that can be literal causes and effects.¹⁷

But neither mind-independence nor naturalness is a requirement for causal efficacy.¹⁸ Many socially constructed entities, like money and art, are taken to be straightforwardly causal. Like social categories, they figure into intuitively true counterfactual causal claims. They back explanations better than relevant alternatives,

¹⁵ See Webster (forthcoming) for an elaboration of this view.

¹⁶ While here I discuss reasons as a vehicle for the causal power of social categories, the view that social categories are causal does not depend on this. The suggestion that social categories restrain agency can be viewed as independent of any claim about reasons.

¹⁷ See Ritchie and Mason (forthcoming) for a discussion of the ontological and causal status of social categories. See Bernstein (2021) for an argument that the most fundamental level of reality need not be the top-most or bottom-most level.

¹⁸ For relevant discussion of the naturalness of social categories, see Mason and Ritchie (forthcoming).

like reductions to lower-level physical causal processes. And the social world itself might turn out to be more fundamental than the metaphysical levels “above” and “below” it.¹⁹

One might try to accept one of a few theoretical alternatives to the causal power of social categories. First, one might be tempted by the idea that being perceived as belonging to a particular social category is what is doing the causal work. For example, when DeShawn is pulled over because he is Black, he is pulled over because he is perceived to be Black. When someone would have had an easier time in a negotiation if she were not a woman, it’s really because she was perceived to be a woman. This view locates causal power of social categories in perceptions of people belonging to those categories.

There are several reasons to avoid such a move. One reason is that perceptions of social categories are not always causally intersubstitutable with social categories. Consider “If Bailey hadn’t been trans, he wouldn’t have struggled with what sort of trousers to wear.” Supposing that Bailey’s trouser dilemmas are private to him and involve his own reasons, external social perception does not play a role in his private deliberations. Another reason is that even in cases where the perceptual claim comes out true, it doesn’t give us the whole explanation. Consider “If Meena hadn’t been a Dalit woman, she would have had different marriage prospects.” Here, we can certainly theorize about what would have happened if Meena hadn’t been perceived as a Dalit woman. But perception doesn’t capture the social history of the category “Dalit woman” that fully explains Meena’s marriage prospects. Social perceptions are not playing the same causal role as the social category.²⁰ As above, it is the category itself, rather than the perception of the category, that generates the informative and correct causal claim. Finally, note that even perception of social categories is causally downstream of the existence of the social category itself. For example, Meena is perceived as a Dalit woman because her perceivers already have concepts associated with the category “Dalit woman”.

One might locate the causal power of the social world in the possession of social properties rather than social categories. For example, having the social property *Black*

¹⁹ See my (2021) for an argument that the “middle” level of reality could be the most fundamental.

²⁰ Thanks to Aness Webster for this point.

causes DeShawn to be pulled over, rather than the social category *Blackness* doing the causal work; and possessing the social property *Dalit woman* influences Meena's marriage prospects, rather than the social category itself.

I do not take this approach to differ strongly from mine. It might turn out that social categories are just mass conferrals of social properties on individuals,²¹ or that social property instances "participate" in broader universals, à la David Armstrong. What is important for my view is that social categories causally influence the lives of their members. For our purposes, the ontological details are less important than the causal ones (though, of course, I also take the former to be of intrinsic metaphysical interest.)

I will not further argue for the causal power of social categories here. Even if one is skeptical of the strong claim, it is enough for our purposes that social categories generate intuitively true causal counterfactuals about human lives, and powerfully back predictions and explanations in such contexts. As I will suggest in the next section, the causal power of social categories results in non-ideal agency for many human individuals.

2. Membership in Social Categories as a Fundamental Restriction on Human Agency

Given that we are unavoidably embedded in a socially categorized world, and that social categories have pervasive causal power in our lives, membership in many social categories constitutes a fundamental restriction on human agency. One cannot choose the categories into which one is born; one cannot escape social categorization; one cannot individually choose which categories exist; and one can only change the natures and causal power of categories with great difficulty. In this section I will elaborate on some of these constraints. Together, I suggest, they create a situation of non-ideal agency for socially embedded humans: it is very difficult to alter one's position in the system, and it is equally difficult to alter the system itself. The agency is non-ideal because humans are working within unchangeable social constraints.

²¹ For this sort of view, see Mallon (2016), according to which social kinds are causal property clusters. Mallon's view is compatible with the sketch of my view given here.

To clarify this claim, a word on how I am using the term “agency” will be helpful.²² In the free will literature, agency is a technical term. Davidson (1963), for example, views agency as involving a particular relationship between agents and events. Ginet (1990) explains agency as involving particular acts of volition. Here I will use the term and the concept in a broader sense to mean something like “human freedom to define one’s life course.”²³ Though I do hold that membership in social categories does sometimes restrict agency in the technical sense (for example, in the sense of restricting control over events, volitions, or reasons-sensitivity), my arguments will revolve around a more expansive understanding of the concept.

Consider that people are born into social categories that shape their social and economic lives. To return to the example above: a Dalit female is born into a particular caste as a particular gender, and these categories jointly influence her marriage and life prospects. In the United States, level of education is the biggest determinant of future earnings, and education levels are stratified by racial category.²⁴ Warren Buffett claimed: “... if I had been a female, my life would have been entirely different. [...] I was born in 1930, I had two sisters that have every bit the intelligence that I had, have every bit the drive, but they didn’t have the same opportunities.”²⁵ No matter how tenacious one is, one cannot escape one’s initial social conditions.

Since one cannot choose one’s initial membership in social categories, a large swath of life experience is caused by social categorization outside of one’s control. Bornstein (1994) describes the situation of gender categorization starkly:

“We’re born: a doctor assigns us a gender. It’s documented by the state, enforced by the legal profession, sanctified by the church, and it’s bought and sold in the media. We have no say in our gender – we’re not allowed to question it, play with

²² Thanks to several NOWAR participants for pressing me on the sense of agency under discussion.

²³ For precedent broadening the understanding of agency beyond its technical sense, see Susan Wolf’s (1987) argument that the metaphysics of moral responsibility leaves out sanity as a requirement on moral responsibility. Wolf seeks to refocus the debate about moral responsibility as involving the everyday concept as entertained by “lawyers, judges, and parents” (1987: p. 1). She does not oppose technical senses of moral responsibility or debates about them; rather, she seeks to analyze a broader, everyday concept of responsibility. I take myself to be doing something similar with agency.

²⁴ <https://www.census.gov/library/publications/2021/demo/p60-273.html>

²⁵ <https://www.cnbc.com/2018/10/16/warren-buffett-says-being-a-white-man-helped-him-succeed.html>

it, work it out with our friends, lovers, or family. Gender is not consensual.”
(1994, p. 123)

While gender is an obvious and timely example of externally imposed social categorization, much social categorization is non-consensual.

Which social categories one is born into is a matter of luck: obviously, one does not choose the social circumstances of one’s entry into the world. I call this *taxonomic luck*. Taxonomic luck describes the fundamental human situation of being unable to choose the social categories that initially causally influence one’s life. This sort of luck can be viewed as a type of Nagelian constitutive moral luck, since it involves luck in who one is, and the traits and dispositions one has.²⁶

Like its genus constitutive luck, taxonomic luck shapes moral responsibility. Consider a person who is born into a minoritized social category and has fewer opportunities to socially oppress other individuals in her life. Her actions might have been different if she had been born into a privileged social category. Someone who is born into a societally dominant social category who participates in social oppression of those in marginalized social categories might not have done so if they were born into different circumstances. Taxonomic luck, then, can be understood as a sort of moral luck: our moral trajectories are heavily shaped by our membership in social categories.²⁷

It is very difficult to exit social categories to which one belongs, and to enter new ones. Below, I discuss cases in which people do try to shift the social categories to which they belong. As I will suggest, shifting one’s membership in social categories does not eliminate their power in human lives: initial social conditions play a large role in one’s lifelong trajectory. Even the attempt to change social categories is causally downstream of one’s initial social conditions.

Such attempts at category change are also constrained by which social categories there are. As Dembroff (2018) suggests, sometimes the very existence of social categories, like binary gender and racial categories, harms and constrains individuals. People who do not cleanly fit into preexisting gender and racial categories are socially

²⁶ See Nagel (1993) for the full description of his four types of moral luck.

²⁷ For a good discussion of how social categories define participation in collective structural injustice, see Zheng (2021).

punished for their non-conformity. People who do fit naturally into such categories, especially socially dominant ones, are elevated into positions of privilege, whether or not they endorse the categorization scheme.²⁸ And people who attempt to refuse categorization within the societally accepted scheme are still categorized, and thus constrained, by others working within the dominant category schema.

Being socially categorized is inescapable. In every social interaction, one's social category influences how one is treated. The effects of this treatment range from small (microaggressions) to cumulatively large (lost wages and opportunities over a lifetime of being in a socially marginalized group).²⁹ Even small children are socially categorized in ways that unjustly constrain their autonomy and agency.³⁰

Social categories are causally rigid. Individuals generally cannot change the causal profiles of social categories, including those to which they belong, though I address such attempts in section 4 below. For example, one person cannot control the fact that the social category *woman* is associated with certain stereotypes that infuse everyday social interactions. Members of minoritized racial groups cannot control patterns of marginalizing treatment based on category membership. Members of marginalized sexual identities cannot control heteronormative projections on their preferences and behavior. Members of intersectional social categories cannot control how different sorts of social oppression interact and intersect. Those belonging to socially dominant social categories cannot control most dimensions of privilege in their lives.

The nature and causal profile of a social category may change while one is a member of it, and such changes are entirely out of individual control. Sometimes, these causal profiles change for the better: people belonging to them have more opportunities for success and societal integration than those who came before them. For example, the nature of the social category *lesbian in the United States* has changed radically from 1970 to now, as has the social category *trans*. But sometimes a change in the causal power of social categories also results in a loss of social power: the causal profile of the category

²⁸ See Manne (2017) for a discussion of how patriarchy elevates and rewards women who conform to patriarchal norms.

²⁹ This inescapability holds for socially privileged positions as well as socially dominant ones. For example, one cannot easily escape one's white privilege when being pulled over by police.

³⁰ See Kukla (2020) for a good discussion of children's autonomy and agency.

landed gentry, for example, has greatly shifted from a generation ago. Muslims in the United States suffered greater discrimination and social oppression after the 9/11 attacks. Being a Black man in the United States has become more dangerous in the past forty years due to police brutality.

Even forms of agency enabled by membership in social categories cannot be chosen. One cannot choose the forms of privilege and special access connected to one's social category membership. A (non-passing) Black woman cannot choose to have the forms of privilege associated with being white; and a straight, white, cisgender man cannot have access to special ways of experiencing being gay, Black, and non-binary. So even though social categories might enable as well as constrain, enablements are subject to the same problem of not being chosen by their bearers.

What would ideal agency look like in a socially categorized world? This is a tricky question, especially given that even the time and place of one's birth is a kind of constraining categorization.³¹ Here's one picture: One is born but not immediately categorized; one makes a fully informed decision about which, if any, social categories to join when one comes of age³²; people perceive one as belonging to the social categories to which one consciously opts in; one can exit social categories and their constraints at will.³³ Social categorization would be consensual, to use a framework suggested by George and Briggs (manuscript). Social categories would not exert downward causal control on one's life. And even if categorization is in some sense unavoidable, social categories would not hold the same moral and practical import that they in fact do.

The actual world is very far from this scenario. We are constrained by social categorization in ways that deeply effect our lives. In what follows, I describe and address different sorts of attempts to resist the causal powers of social categories. We can exercise some agency with respect to social categories, I suggest, but ultimately such forms of resistance have limited power in shaping the vast causal influence of social categories.

³¹ Thanks to David Shoemaker for this point.

³² Veronica Roth's 2011 science fiction novel *Divergent* explores this scenario, though the characters are still heavily constrained once they join the social categories that they have chosen. See also George and Briggs (manuscript, p. 6) for discussion of a more ideal categorized world.

³³ Changing the causal profiles of social categories wouldn't be as important to agency in such a scenario, since one could enter and exit categories at will.

3. Forms of Social Category Resistance

For each way that social categories constrain human agency, there is a corresponding way to resist each constraint. People may attempt to shift social category membership by intentionally exiting or entering social categories; by creating, destroying, or altering social categories; or attempting to change the causal profile of social categories. Each form of resistance is highly limited, however, and humans are still forced to operate in a situation of non-ideal human agency. In this section I will elaborate on these claims.

Consider altering one's membership in social categories. First, note that it is extremely difficult to enter and exit social categories, since doing so involves not just a realigning of one's "felt" categories, but is also a matter of how one is perceived. One common such alteration in social category membership is a social, legal, and/ or medical change in gender identity, especially as undertaken by trans individuals.³⁴ Trans identities embody several different forms of social category resistance, including shifts in social category membership, attempts to change the social power of the categories in one's life, and (on some occasions) aiming to change the very natures of gender concepts in inhabiting a gender identity. In many cases, people seek to change social categories not only to affirm the gender identity they presently feel or have always felt, but to change the way they are perceived by others.

In changing one's membership in a social category, one thereby changes the causal profile of the category in one's life.³⁵ But as many will attest, such changes alter the causal constraints and enablements imposed by gender categories rather than eliminating them. Some who undergo medical transition and who pass under their new gender identity in everyday society are faced with the causal power of the new social category. Neuroscientist Ben Barnes marveled that he "was able to complete a whole sentence without being interrupted by a man. A colleague who didn't know he was

³⁴ Here I focus on cases in which people seek the gender identity to which they are perceived to belong, but this does not describe every case. See Rachlin (2018) for a discussion of medical transitions that are not intended to be social transitions.

³⁵ While here I primarily focus on ways that social categories restrict individuals, social categories can and do enable new opportunities and privileges as well. In rural Columbia, for example, local politicians have historically altered their own membership in indigenous groups in order to be elected and able to govern certain local regions. Thanks to Santiago Amaya for this example.

transgender even praised his work as “much better than his sister’s.”³⁶ Those happy to adopt the identity of trans men or trans women nonetheless face the new causal constraints and enablements imposed by trans social categories. Transitioning can also bring new forms of intersectional oppression. Trystan Cotton, a post-transition Black man, explains:

“I get pulled over a lot more now. I got pulled over more in the first two years after my transition than I did the entire 20 years I was driving before that. Before, when I’d been stopped even for real violations like driving 100 miles an hour, I got off. In fact, when it happened in Atlanta the officer and I got into a great conversation about the Braves. Now the first two questions they ask are: Do I have any weapons in the car, and am I on parole or probation? [...] Being a black man has changed the way I move in the world. I used to walk quickly or run to catch a bus. Now I walk at a slower pace, and if I’m late I don’t dare rush. I am hyper-aware of making sudden or abrupt movements, especially in airports, train stations and other public places. I avoid engaging with unfamiliar white folks, especially white women. If they catch my eye, white women usually clutch their purses and cross the street. While I love urban aesthetics, I stopped wearing hoodies and traded my baggy jeans, oversized jerseys and colorful skullcaps for closefitting jeans, khakis and sweaters. These changes blunt assumptions that I’m going to snatch purses or merchandise, or jump the subway turnstile. The less visible I am, the better my chances of surviving.”³⁷

While one can exit the gender category into which one is born and choose a new category, one is never entirely free of related social constraints and enablements, since the causal role of the new category replaces the old one.

Exiting social categories is additionally difficult because doing so can be more than a matter of appearance and bodily change. Some religious social categories are very difficult to exit. Disavowing particular religious beliefs is neither necessary nor sufficient

³⁶ <https://www.washingtonpost.com/news/local/wp/2018/07/20/feature/crossing-the-divide-do-men-really-have-it-easier-these-transgender-guys-found-the-truth-was-more-complex/>

³⁷ Ibid.

in some cases of religious categorization, especially those involving “cultural” religiousness. Friedell (manuscript) takes up the case “in which someone has been a member of the Jewish community but simply wants to stop being a member, without converting to any other religion.” (p. 10) On the puzzle of what it would take to make himself non-Jewish, he writes:

“Something changed in graduate school. I became disillusioned with religion. Not theology but religion as a social category. What once struck me as a source of community now seemed to be a source of division [...] I was no longer satisfied to self-identify, as many American Jews do, as “culturally” but not “religiously” Jewish. That way of identifying puzzled me. What does it mean to be “culturally” Jewish, other than to embody or embrace certain stereotypes, such as being neurotic and enjoying *Seinfeld* reruns? Even worse, this common way of identifying still seemed to indirectly support the divisiveness of religion as a social category. Why not go a step further and identify as non-Jewish? [...] I declared to some friends that I was no longer Jewish. I wasn’t converting to another religion. I wasn’t intending to conceal my Jewish upbringing. Nor did I plan to stop being neurotic or to stop enjoying *Seinfeld* reruns. I merely identified as non-Jewish. I immediately questioned whether such a change was possible. Can a Jewish person become non-Jewish? Or is a Jewish person always Jewish?” (manuscript, p. 3)

What is notable about the case is that Friedell has difficulty exiting Jewishness *qua* social category rather than *qua* religious category. He argues that it is very challenging to become non-Jewish without converting one’s religious beliefs, given the interplay of the religious category and the social category. Even disavowing religious beliefs is not enough in some cases. Catholics seeking to become non-Catholic face a similar problem: short of religious conversion, excommunication is the major route to becoming non-Catholic.³⁸

³⁸ <https://www.theguardian.com/commentisfree/2020/apr/11/how-do-you-become-formally-not-a-catholic-you-take-the-law-into-your-own-hands>

A related form of social category resistance is the alteration of the causal profile of a social category—the attempt to change how it behaves in society and thus in one’s own life. To change the causal profile of a social category is to alter the social reasons furnished by the category. On a simple level, consciously defying a stereotype is a way of trying to change the power of a social category. Since the stereotype influences how members of a social category are treated, defying a stereotype changes the mediating social reasons furnished by the social category. For example, in defying a stereotype of womanhood by being assertive in the workplace, one replaces one set of social reasons (“She is quiet in the workplace because she is a woman”) with another (“She is not a ‘typical’ woman because she is assertive in the workplace”).

However, like entering and exiting social categories, altering the causal profiles of social categories is extremely difficult. Individuals generally do not have power over entire social categories. A private individual defying a stereotype about womanhood, for example, can at most result in changing a few people’s minds about the category *womanhood*. Changing the social reasons to which people respond is generally done on a small scale. Moreover, those attempting to change the causal profiles of social categories are forced to make tradeoffs with respect to which aspects they alter. Webster writes of this phenomenon:

“Especially those who are members of more than one marginalized group may bargain with various social norms to which they feel a certain sort of pull. For example, a black woman might refrain from straightening her curly hair although she continues to wear make-up. Moreover, for many, this decision is related: the costs of violating one norm of feminine appearance are placated by her choice to comply with some another norm. To give another example, a woman, animated by gender equality, might be very happy to be vocal about policies affecting gender equality at faculty meetings, but might be quiet about other issues that also matter to her. Similarly, one might attempt to balance or outweigh the potential costs or downstream of not fitting into the stereotype—say, of being submissive or subdued which is a norm of Asian femininity—by engaging in various actions to offset those costs—by being friendly, by going out of one’s way not to

inconvenience people, and by conforming to certain other norms of feminine appearance and presentation.” (2021, p. 115)

One notable feature of this example is that the causation moves in both directions between the agent and the social category: the woman in this example attempts to influence the nature of the social category, but her efforts are also thereby influenced by the rigidity of the social category.³⁹

Social taxonomies are not morally neutral: how categories are carved, and which categories exist, reflect societal attitudes and biases. Thus another form of resistance is to shape the very existence and conceptual content of social categories themselves. A recent methodological movement in social philosophy seeks to *amelioriate* social concepts. Roughly, ameliorative analyses seek to shape concepts into ones that we should be using rather than the ones that we actually use. Social justice is a central motivation in evaluating what concepts should be used, and hence in ameliorative projects.

Amelioration is a type of *conceptual engineering*, the more general project of changing the nature of concepts. Both amelioration and conceptual engineering can be viewed as attempts to directly causally intervene on the nature of social categories by changing them.⁴⁰ Ameliorating the concept of marriage to include same-sex marriage, for example, changes the causal power of the concept: same-sex couples are granted the same legal status as other couples. Ameliorating the concept of “parent” to include adoptive parents changes the causal profile of the social category *parent*.

As Bettcher (2014) notes, the very existence of trans identities is ameliorative. She writes:

“In a beyond-the-binary model, to say that trans people are marginal with respect to the binary is to locate them in terms of the categories “man” and “woman” as dominantly understood. If trans bodies can have different resistant meanings, the

³⁹ This is further evidence there is no clean distinction to be had between internal and external social categories: the two types are causally interactive, rendering the distinction porous.

⁴⁰ Famously, Haslanger (2000) seeks to ameliorate the concept of gender. Díaz-León (2016) suggests that ‘woman’ is a politically significant term. Manne (2017) applies an ameliorative analysis to the concept of misogyny. A德勒 (forthcoming) gives an ameliorative account of sexual orientation categories. Bettcher (2014) gives an ameliorative analysis of gender terms as applied to trans and nonbinary individuals.

decision to say of those bodies that they are “mixed” or “in between” is precisely to assume a dominant interpretation. So the problem is not the rigidity of the binary categories but rather the starting assumption that there is only one interpretation in the first place-- the dominant one. Similarly, in the wrong-body model, to become a woman or a man requires genital reconstruction surgery as the correction of wrongness. But this is to accept a dominant understanding of what a man or a woman is.” (2014, p. 390)

On Bettcher’s view, trans bodies and trans identities resist the gender binary both by transcending it, and by embodying multiple interpretations of it.

Ameliorative social category resistance is not limited to the existing social taxonomy. Call *ameliorative creation* the process of creating new social categories for social justice-oriented ends, and call *ameliorative destruction* the process of destroying existing social categories for social justice-oriented ends. The widespread adoption of the category *nonbinary* is an example of ameliorative creation: it adds a category for individuals who do not conform to traditional gender categories. Another example: Andler (forthcoming) argues for a social categorization scheme that includes queerness, since people who are queer face specific forms of heteronormative oppression. Such a category is not accounted for by the traditional distinction between gender or sexual orientation. In contrast, abolitionists about gender aim for ameliorative destruction.⁴¹ There are also abolitionist perspectives on race, religious categories, mental health categories, and citizenship categories.

But amelioration also faces obstacles as a method of reclamation of social agency. First, it requires widespread adoption of the ameliorated concept or concepts. As Podonsky (forthcoming) notes, an effective ameliorative project should change the functions of the social categories in everyday society. It is not enough to change the concept in the philosophical community; the changed concept should result in widespread differences in social and legal practice. Localized ameliorated categories are generally not enough for widespread social change. Individuals generally do not have the power to

⁴¹ See Cull (2019) for an extended discussion of gender abolitionism. They argue against the view on the grounds that it does not respect trans rights.

ameliorate concepts that matter to them. One can choose to be treated as having a nonbinary gender classification, but one cannot choose to be treated as having a cactus gender classification beyond a small set of accommodating acquaintances.⁴²

Second: in order to be effective, ameliorators must reach those who are not sympathetic to the aims of amelioration. Saul (2012) explains that “feminists need to communicate successfully both with each other and with those who are not (yet) feminists, feminists should want to avoid large-scale misunderstandings wherever possible.” (2012, p. 197) Such attempts have, in the past, been successful: 70% of Americans now support same-sex marriage,⁴³ a dramatic shift in attitudes about the concept of marriage, among other things. But these shifts are rare and often take several generations.

Third, amelioration risks what Marques (2020) calls *meaning perversion*, which occurs when a concept is adopted or manipulated for bad ends. When people try to deny that immigrants are truly American, for example, they are attempting to ameliorate a concept in dangerous ways.⁴⁴ Ameliorating the concept of pornography to include any image of women would express a particular view about women’s bodies and value. Amelioration can, and does, run amok. It is subject to the same historical forces as any other social justice project. Amelioration can be morally hazardous if undertaken in the wrong way, in the wrong era, or in the wrong social context.

4. Applications

These results shed new light on three timely topics in the philosophy of gender: gender pronouns, the rights of trans and nonbinary individuals, and cases of so-called transracialism. Each case can be understood as involving social category resistance.

Conversational declaration of one’s gender pronouns (“she/ her” or “they/ them,” for example) has become commonplace. In addition to the straightforward function of the declarations—to notify people of one’s gender identity—such declarations are also a

⁴² See Dillon (2007) for an argument that their gender is a cactus.

⁴³ <https://news.gallup.com/poll/350486/record-high-support-same-sex-marriage.aspx>

⁴⁴ Danger is in the eye of the beholder. Some on the political right hold that ameliorating the concept of gender is dangerous. Thanks to David Shoemaker for this point.

form of social category resistance. Pronoun declarers resist nonconsensual gendering by proactively declaring their social category membership. Such declarations are intended to mediate social responses. The act of pronoun declaration shapes the causal constraints imposed by social categories, since it expands respect-based social reasons to which others should respond.⁴⁵ To put it in terms of Austin's speech act theory: the perlocutionary force of pronoun declaration is to notify others of one's gender identity, while an illocutionary force of the declaration is to shape how the social category behaves.⁴⁶ Respecting declared gender pronouns, particularly for trans and non-binary individuals, is also a form of social category resistance.⁴⁷ On the flipside, intentionally misgendering people functions as a form of rigid social category enforcement.⁴⁸

Relatedly, one among many reasons to uphold trans rights is that trans individuals are reclaiming agency in a world of nonconsensual social categorization. The right to categorize oneself lies at the philosophical foundation of trans rights. Opting in to social categories is a way to exercise non-ideal social agency. This framework comports with contemporary views of trans identities: Bettcher, for example, "[understands] a trans world in resistant relation to a dominant world." (2014, p. 389) Bell (forthcoming) argues that trans conformity to gender norms is a way of enacting personal authenticity in a non-ideal world. They "[understand] authenticity as a socially embedded, constructive project undertaken in a non-ideal social world." (forthcoming, p. 1)

Not all cases of social category resistance are respected and encouraged. In a widely publicized case, Rachel Dolezal, a white woman from the Pacific Northwest of the United States, pretended that she was Black for many years, accruing trust and resources from the Black community and leading the local chapter of the NAACP. In a 2015 interview, Dolezal claimed: "[I] self-identified with the black experience. I was

⁴⁵ See George and Briggs (manuscript) for a discussion of what would be required for genuinely consensual gendering.

⁴⁶ See Hernandez (forthcoming) for an argument that correctly gendering people is a form of loving attention.

⁴⁷ Gender pronoun declarations by people who pass as straight and cisgender can also function as social category resistance: they challenge assumptions about what sort of people make such declarations. The adoption of "they/them" by people who identify as cis is an additional form of resistance. Dembroff and Wodak (2018) suggest that all people should adopt gender-neutral pronouns, partly in solidarity with those who do not conform to the binary.

⁴⁸ See Kapusta (2016) for a description of misgendering as a moral harm against trans individuals.

drawing self portraits with the black crayon instead of the peach crayon.”⁴⁹ A significant controversy ensued, both over Dolezal’s behaviors and over the philosophical issues raised by the situation of so-called “transracialism.”⁵⁰

Among other things, transracialism endorses the idea that self-identification is sufficient for membership in a social category. Transracialism can be understood as a form of morally problematic social category resistance, though the differences between these cases and broadly parallel ones are hotly disputed.⁵¹ Such cases also reveal contemporary moral attitudes about the sanctity and inviolability of certain social categories. As Dembroff and Payton (2018) note, even these attitudes are conventional and tied to each historical moment. It is possible transracialism will be differently morally considered in a future, less racially stratified world. What is objectionable about transracialism stems, in part, from pervasive racial injustice. The social and the moral are intertwined: in a more equal world, perhaps self-identifying with racial categories would not be morally fraught.⁵²

5. Conclusion

This paper has argued that social categorization creates a situation of non-ideal agency in human life: social categories are imposed on us from birth, and their causal power pervades the human condition. Resisting social categories takes several forms, but does not ever result in total freedom from taxonomic constraints. Rousseau famously notes that we are born free, but everywhere we are in chains.⁵³ In fact, it is the opposite: we are born constrained and categorized, and we cast off the chains only when we become aware of our non-ideal social conditions.⁵⁴

⁴⁹ <https://www.refinery29.com/en-us/2015/06/89156/rachel-dolezal-transracial>

⁵⁰ See Tuel (2017) for a controversial take on such cases.

⁵¹ See Dembroff and Payton (2018) for an argument for the moral asymmetry of race and gender in these cases.

⁵² It might turn out, however, that race is such a distinctive social category that the best form of resistance is eliminativism. Thanks to Aness Webster for this way of thinking about it.

⁵³ This is the famous first line of *The Social Contract* (1762).

⁵⁴ I owe thanks to Santiago Amaya, Oisín Deery, Daniel Nolan, Michael Rea, David Shoemaker, Manuel Vargas, Aness Webster, two anonymous referees, and the participants in the New Orleans Workshop of Agency and Responsibility for valuable feedback on this paper. Thanks also to the National Endowment for the Humanities: revisions on this paper were completed during my fellowship year.

References

- Andler, Matt (forthcoming) “Public Health, Political Solidarity, and the Ethics of Orientation Ascriptions” *Ergo*.
- Ásta (2018) *Categories We Live By: The Construction of Sex, Gender, Race, and Other Social Categories* New York, USA: Oxford University Press.
- Bell, Rowan (forthcoming) “Being Your Best Self: Authenticity, Morality, and Gender Norms” *Hypatia*.
- Bernstein, Sara (2020) “The metaphysics of intersectionality” *Philosophical Studies* 177 (2):321-335.
- Bernstein, Sara (2021) “Could a middle level be the most fundamental?” *Philosophical Studies* 178 (4):1065-1078.
- Bernstein, Sara. (manuscript) “Fundamental Social Causes”.
- Bettcher, Talia Mae (2014) “Trapped in the Wrong Theory: Re-Thinking Trans Oppression and Resistance” *Signs* 39 (2):383-406.
- Bird, Alexander (1998) “Dispositions and Antidotes” *The Philosophical Quarterly* 48: 227–234.
- Bornstein, Kate (1994) *Gender Outlaw*. New York: Routledge.
- Cull, Matthew J. (2019) “Against Abolition” *Feminist Philosophy Quarterly* 5 (3).
- Davidson, Donald (1963) “Actions, Reasons, and Causes”, reprinted in *Essays on Actions and Events* (1980), Oxford: Clarendon Press.
- Dembroff, Robin & Wodak, Daniel (2018) “He/She/They/Ze.” *Ergo: An Open Access Journal of Philosophy* 5.
- Dembroff, Robin (2018) “Real Talk on the Metaphysics of Gender” *Philosophical Topics* 46 (2):21-50.
- Dembroff, Robin & Saint-Croix, Cat (2019). ‘Yep, I’m Gay’: Understanding Agential Identity. *Ergo: An Open Access Journal of Philosophy* 6:571-599.
- Dembroff, Robin and Payon, Dee (2020) “Why We Shouldn’t Compare Transracial to Transgender Identity” *Boston Review*.
- Díaz-León, Esa (2016) “‘Woman’ as a Politically Significant Term: A Solution to the Puzzle” *Hypatia* 31(2), pp. 245-58.

Dillon, Rax Espeon (2007) “My gender is cactus” URL <http://rax.livejournal.com/104836.html>.

DuBois, W.E.B. (1903) <https://www.gutenberg.org/files/408/408-h/408-h.htm>

Fanon, Frantz. (1952) *Black skin, white masks*. Grove press. (revised 2008 version)

Friedell, David. (manuscript) “Becoming Non-Jewish”

George, B. R. & Briggs, R. A. (manuscript) “Science Fiction Double Feature: Trans Liberation on Twin Earth”

Ginet, Carl (1990) *On Action*, Cambridge: Cambridge University Press.

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

Hill Collins, Patricia. (1990) *Black Feminist Thought*. New York: Routledge Press.

Hernandez, Em (forthcoming) “Gender Affirmation and Loving Attention” *Hypatia*.

Jenkins, Katharine (2020) “Ontic Injustice” *Journal of the American Philosophical Association*. 6 (2):188-205.

Kapusta, Stephanie (2016) “Misgendering and its moral contestability.” *Hypatia* 31 (3): 502-519.

Kukla, Quill (2020) “Taking Children’s Autonomy Seriously as a Parent.” *APA Newsletter on Feminism and Philosophy*.

Mallon, Ron (2016) *The Construction of Human Kinds*. New York: Oxford University Press.

Manne, Kate (2017) *Down Girl: The Logic of Misogyny*. New York: Oxford University Press.

Marques, Teresa (2020) “Amelioration vs. Perversion in Shifting Concepts: The Philosophy and Psychology of Conceptual Variability” Teresa Marques and Asa Wikforss (eds.). pp. 260 – 284. Oxford: Oxford University Press.

Mason, Rebecca & Ritchie, Katherine (forthcoming) “Social Ontology” In Ricki Bliss & James Miller (eds.), *Routledge Handbook of Metametaphysics*.

McKittrick, Jennifer (2004) “A Defence of the Causal Efficacy of Dispositions” *Sats: Nordic Journal of Philosophy*, 5: 110–130.

- Nagel, Thomas (1993) "Moral Luck" in *Moral Luck*. Daniel Statman (Ed.). State University of New York Press, Albany, New York, pp. 57-71.
- Passinsky, Asya (2020). Should Bitcoin Be Classified as Money? *Journal of Social Ontology* 6 (2):281-292.
- Podonsky, Paul-Mikhail (forthcoming) "Can Social Engineering Actually Promote Social Justice?" *Synthese*.
- Rachlin, K "Medical Transition without Social Transition: Expanding Options for Privately Gendered Bodies" (2018) *Transgender Studies Quarterly* 5 (2): 228–244.
- Rousseau, Jean-Jacques (1947) "The Social Contract" (ed. Charles Frankel) New York: Hafner Publishing Co.
- Tuvel, Rebecca (2017) "In Defense of Transracialism" *Hypatia* 32 (2):263-278.
- Webster, Aness (forthcoming) "Socially-Embedded Agency: Lessons from Marginalised Identities" In David Shoemaker (ed.), *Oxford Studies in Agency and Responsibility* 7.
- Wolf, Susan (1987) Sanity and the Metaphysics of Responsibility. In Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press. pp. 46-62.
- Zheng, Robin. (2021) "Moral Criticism and Structural Injustice" *Mind* 130 (518):503-535.