

Social Dependence

Sara Bernstein
University of California, Santa Cruz

(forthcoming in *Australasian Journal of Philosophy*; please cite published version)

Abstract: We often reason about what our lives would have been like if we had belonged to different social groups— for example, ‘If I had been Black, being pulled over by police would have been more frightening’, or ‘If I had not been a woman, I would have had an easier time in that meeting’. This paper makes sense of such *countersocial counterfactuals*, conditionals whose antecedents run contrary to social facts, and in many cases, contrary to identity facts and essentiality facts. It is suggested that many countersocials are counterpossibles. Based on the non-trivial truth of some countersocials, it is argued that social categories are literally causal: they are causes and effects, both of other general phenomena and some particular events in our lives. These threads are then brought together in order to examine intersectional oppression. Intersectional oppression can be illuminated by appeal to proportionate causation, or causation containing the appropriate level of causal detail.

Key words: counterfactuals, counterpossibles, feminist philosophy, philosophy of race, intersectionality, proportionality

‘I wouldn’t have had the life I’ve had if I hadn’t been gay.’—Elton John

‘[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.’—Warren Buffett

‘I think if I hadn’t been black, I wouldn’t have been as popular as I was. I would have just done whatever the majority was doing.’—Tycely Williams

1. Introduction

Like Elton John, Warren Buffett, and Tycely Williams, we often imagine what our lives, personalities, and selves would have been like if we had belonged to different social groups and social categories.¹ Counterfactual reasoning like ‘If I had been Black, being pulled over by police would have been more frightening’, or ‘If I had been a man, I would have had an easier time in

¹ For the initial quotes in context, see Malkin (2019) about Elton John, Elkins (2018) about Warren Buffett, and Colby (2012: 28) about Williams.

that meeting' is a commonplace, widespread psychological and social practice. Many such *countersocial counterfactuals*, counterfactual conditionals whose antecedents run contrary to social fact, play a central role in imagining how our lives and our social world could have been different in ways large and small. Countersocial counterfactuals are central to many kinds of theorising, including psychology, gender studies, ethical theory, economics, and the law.² Key concepts in feminist philosophy, such as discrimination and intersectional oppression, are phenomena that are best understood through countersocial reasoning. But despite their centrality to these types of theorising, relatively little attention has been paid to countersocial counterfactuals in the philosophical literature.

The primary goal of this discussion is to examine and understand countersocial counterfactuals, and how they generate true causal claims involving social categories. The secondary goal is to bridge work on causation and counterfactuals with central topics in the philosophy of gender, the philosophy of race, and social metaphysics more generally. Here's the plan. In section 2, I explore what kinds of countersocials there are, and distinguish between various kinds. In section 3, I suggest that many countersocials are also *counterpossibles*, counterfactuals whose antecedents are impossible. While counterpossibles in the realms of mathematics, logic, and the philosophy of science have been widely explored (Lewis 1973b; Williamson 2007: 156–72; Berto, French *et al.* 2018; French, Girard *et al.* 2022), counterpossibles in social philosophy have gone relatively under-studied. Which countersocials are counterpossibles is a deep and philosophically interesting question, and I discuss why that is. In section 4, I argue that the truth and non-vacuity of countersocials suggests that social categories are causal: social categories can be causes, effects, and causal intermediaries. In section 5, I discuss the explanatory indispensability of impossibility in understanding countersocial counterfactuals, and conclude that impossible worlds play an important role in an account of social causation. In section 6, I bring the threads together in order to apply the results to the phenomenon of intersectional oppression. I suggest that intersectional oppression can be illuminated by appeal to proportionate causation, causation containing the appropriate level of causal detail.

² For examples of countersocials in psychology, see Epstude and Jonas (2015); for gender studies, see Shapiro (2007); for ethical theory, see Appiah (1990); for economics, see Pesaran & Smith (2016); and for law, see Levy and Ben-David (2015).

2. Different Sorts of Countersocials

Countersocials are counterfactual conditionals that run contrary to social fact.³ In this discussion I will focus primarily on countersocials that involve social categories, their natures, and membership in them, as a way to open the investigation into countersocials. But there are many more sorts of countersocials that are worth investigating.

For taxonomic purposes, distinguish between *local* and *global* countersocials. A local countersocial concerns a token situation. The countersocials with which we began, ‘If I had been Black, being pulled over by police would have been more frightening’ and ‘If I had not been a woman, I would have had an easier time in that meeting’ are local: they range over specific, discrete situations. Global countersocials, in contrast, have an extended range. Consider: ‘If I had been Black, I would not have had the same professional opportunities’ and ‘If I had been a trans man, my friends would have been supportive’. In these cases, the consequent runs contrary to a wide, diffuse portion of a person’s personal history: it is not just a particular situation that would have been different, but large swaths of the person’s life. Global countersocials encompass many situations. In the following discussion, I will talk about both.

While some countersocials have wide consequents, other countersocials additionally have wide antecedents. Consider ‘If Mexican-Americans were treated like French-Americans, the city of El Paso, Texas would be wealthier’. Both the antecedent and the consequent are broad. The idea here is that if the treatment of Mexican-Americans, as a broad social category, was more like another social category, a city with a large population of people belonging to the category would be much different. Consider also ‘If ‘Don’t Ask, Don’t Tell’ hadn’t been repealed, the experience of LGBTQ people in the military would have been different’. This countersocial involves a large, diffuse consequent: treatment of people who are gay across a broad range of time in the military. Similarly with ‘If there were more women than men on the US Supreme Court, the interpretation of the laws would be very different’. Global countersocials are frequently used in historical theorising (Nolan 2013).

Counterfactual evaluation is a matter of holding certain things fixed while varying others. In evaluating a typical counterfactual like ‘If my alarm had not sounded, I would not have made

³ For recent discussion about what makes something a distinctively social property, see Mallon (2016), Payton (2022), Pagano (2026).

the train’, it is obvious what we hold fixed and what we vary: we hold fixed everything leading up to the sounding of the alarm, hole-punch out the single event of the alarm sounding and replace it with an alarm that does not sound, and see which proximate, salient events vary.

While we easily entertain localised countersocials like ‘Would he have said that to me if I were a man?’, the reasoning behind such counterfactuals is subtle. Most social categories are not just labels for physical traits: they are undergirded by conceptually rich ideologies and stereotypes. Consider that, in his 2008 presidential election bid, Barack Obama was commonly called ‘articulate’. In counterfactually evaluating this example (‘Would he have been called ‘articulate’ if he hadn’t been Black?’), it is not enough to suppose that everything is the same but that counterpart Barack Obama just has less melanin in that particular local instance. This counterfactual evaluation holds too much fixed, ignoring the details of social constructions of race, and its associated ideologies and historical antecedents. Rather, the conceptually rich possible world that we entertain is the one that varies Barack Obama’s name, mode of presentation, and his political roots on the South Side of Chicago. We somehow imagine that he inhabits not just a different physical form, but that he inhabits a social category with very different extrinsic features and societal relationships. Similarly for many other such counterracial and countergender counterfactuals, which are not just a matter of swapping out physical traits or parts, but rather a matter of varying historically contingent relationships to society. See Bernstein (2024) for a more detailed discussion of this sort of discriminatory speech evaluated countersocially.

Further questions arise once we are trying to understand these countersocials from the points of view of specific accounts of what it is to belong to a social category. One natural way to test analyses of social category membership is to test their intersubstitutability in countersocial claims. On Ásta’s (2018) conferallist model of social category membership, for example, what it is for a person to be a member of a social category is to have the category socially conferred by the right sort of people in a particular social context. And on Rea’s (2022) view, what it is to have a certain gender is to confer that gender on oneself. Thus ‘If Barack Obama hadn’t been Black’ is further interpreted as ‘If Barack Obama hadn’t had Blackness conferred on him...’ and ‘If I hadn’t been a woman...’ becomes ‘If I hadn’t conferred the property of womanhood on myself...’ But such specifications are not necessarily intersubstitutable *salva veritate* in all countersocial claims. For example, the countersocial ‘If I hadn’t been a woman, I wouldn’t have

been talked over in that meeting' and 'If I hadn't conferred the property of womanhood on myself, I wouldn't have been talked over in that meeting' plausibly have different truth values. This substitution failure might be evidence against particular analyses of social category membership. Due to these intricacies and differences between views, I will keep my discussion as neutral as possible about what it takes to be a member of a social category.

Countersocials pose several other distinctive sorts of challenges to traditional modes of counterfactual evaluation. The first sort of challenge is that countersocials might not obey the traditional model of closeness of worlds for counterfactual evaluation.⁴ According to that account, a counterfactual is true just in case all of the closest worlds in which the antecedent is true are worlds at which the consequent is true. Closeness is to be judged according to a particular metric of similarity. It is of the first importance to avoid big, widespread, diverse violations of natural law; it is of second importance to maximise perfect match of spacetime regions; it is of third importance to avoid small, local violations of law (miracles); and it is of little importance to secure approximate match. (Lewis 1979: 47–48)

It is unclear whether these rules for measuring distance from social actuality are the same as measuring distance from physical actuality. Though the two are in some sense related,⁵ what one varies when entertaining countersocials is different from what one varies in much-discussed causation-related counterfactuals. For example, 'If I hadn't set my alarm clock, I would not have woken up on time' requires different sorts of departures from actuality than 'If men could menstruate, they would brag about how long and how much.'⁶ In the latter sort of case, an appeal to merely physical features of reality, while relevant, is not the most informative explanation of the relevant social phenomenon. Toggling the setting on an alarm clock, and then predicting what happens because of this change, is very different than toggling widespread social constructions. The canonical similarity metric used for other sorts of counterfactuals may not entirely capture these changes.

Different kinds of antecedents introduce complexities into countersocial evaluation. One can vary the membership conditions of a social group, vary features of a person, or vary intrinsic properties of an individual, to name a few different varieties of countersocial antecedent. One can also vary the existence

⁴ Here I assume that something like this account is true for normal counterfactuals, despite well-known problems with it.

⁵ For example, supervenience between physical actuality and social actuality might hold.

⁶ This example is drawn from Gloria Steinem's famous (1983) satirical column on the topic.

or nonexistence of the social categories altogether. ‘If it hadn’t been so bad to be gay in Kentucky, I wouldn’t have left my hometown’ alters local social facts, whereas ‘If the category *nonbinary* had existed in my youth, that is how I would have identified’ alters the very existence of a social category. Like more typical counterfactuals, both speaker intention and context will play a role in what to hold fixed and what to vary. But what will vary will often be social facts of varying scope rather than different localised physical facts.

Another complication is that the natures of social categories are themselves contingent and dependent on collective societal treatment and interaction, so it is an intricate matter which aspects of the categories to hold fixed and which ones to vary. Science fictional scenarios have explored worlds in which women are socially dominant, for example, and worlds in which racial categories are not as socially joint-carving as other sorts of social categories.⁷ Veronica Roth’s *Divergent* explores a post-apocalyptic world in which people choose one of five social functions to which to dedicate themselves for life—creating a society carved at different social joints—but in which some social treatment of these skill sets are retained from the actual world. In these cases, some aspects of the social categories are preserved while others are shifted.

The second sort of challenge is that many countersocials involve appeals to metaphysically impossible worlds as well as possible worlds. I turn now to this topic.

3. Countersocials as True, Non-Vacuous Counterpossibles

Many countersocials are *counterpossibles*,⁸ counterfactuals whose antecedents are metaphysically impossible. Canonical examples of counterpossibles include ‘If Hobbes had squared the circle, small children wouldn’t have cared’ and ‘If two plus two had equaled 5, children’s mathematics textbooks would be different’. Other typically discussed counterpossibles appeal to violations in logic or mathematics (‘If two plus two had not equaled four, the world would have been different’), and are thus relegated to specialised philosophical debate. But almost everyone entertains countersocial counterpossibles: they are an everyday form of

⁷ See, for example, Ursula Le Guin’s *The Left Hand of Darkness* (1969), Octavia Butler’s *Parable of the Sower* (1993), and Naomi Alderman’s *The Power* (2017) for science fictional speculation about the natures of social groups.

⁸ Examples of countersocials that are not counterpossibles include ‘If I had grown up in France, I would have studied at the Sorbonne’ and ‘If I had been more like other Millennials, I would have understood Snapchat’. These are departures from social actuality, but appeal to possible, rather than impossible, worlds.

counterpossible reasoning.

Whether counterpossibles can be true and non-vacuous is a matter of controversy. Many philosophers including Lewis (1973) and Williamson (2007) hold that all counterpossibles are vacuously true owing to the difficulty of assessing impossible antecedents. Lewis's thinking on the matter is that anything goes at impossible worlds: if two plus two had equaled five, who's to say what that world would otherwise be like? Contrarily, others hold that counterpossibles can be non-vacuous owing to the intuitive non-vacuous truth of some key examples.⁹ For example, 'If Professor Smith hadn't failed to prove that $2+2=5$, she would have received tenure and academic fame' seems obviously true. So does 'If I had proved a true contradiction, my audience would have been aghast'. 'If a wooden Penrose triangle had been placed on my desk, my hamburger would have tasted better' seems false. Apparent differences in truth and falsity suggest that not all counterpossibles are vacuously true.

Some countersocials constitute clear examples of non-vacuous counterpossibles that can differ in their truth values. Countersocials thus produce further pressure to have an account of counterpossibles as non-trivial and non-vacuous. Consider the difference between:

(1) 'If she had not been Black, she would not have been pulled over by the police.'

and

(2) 'If she had not been Black, the tulips near Lakeshore Drive would have bloomed more quickly.'

Intuitively, (1) is true and (2) is false. Put another way: manipulating a race variable changes the outcome in the first case, but not in the second case.

Typical counterpossibles in the literature appeal to worlds with different logical or mathematical truths. Varying social categories is in some ways a more complicated pursuit than varying logical or mathematical truths.¹⁰ It is controversial whether social category membership

⁹ For a good overview of compromise positions between vacuism and non-vacuism, see Kocurek (2021).

¹⁰ It is a complex matter which sorts of properties are essential to each individual person, or to all persons. Origin essentialists, for example, hold that anyone would have been a numerically different person had they come from a different egg and sperm. The thinking is that they would not have had the same essential biological properties that

is, or can be, essential to individuals.¹¹ Consider the question of whether gender (construed as distinct from the identity one is assigned at birth based on manifest reproductive organs) is essential to a particular person. It seems essential to Donald Trump, for example, that he is male. But presumably there are people for whom biological sex and gender are not so central to their identities, personalities, outlooks, and life experiences. Philosophers of gender and race are split on the question of the essentiality of gender and race. Witt (2011: xi) suggests that it is a widespread intuition that gender is essential to individuals, and Stoljar (2018) holds that gender is essential to one's self-conception, however that connects up with one's personal identity. Appiah (1990: 495) holds that gender is essential to one's identity, but race is not. Appiah distinguishes between a metaphysical self and an ethical self, and argues that race is central (but perhaps not essential) to the latter.

Whether or not there is an answer to the general question of which social categories are essential to individuals, intuitive variation is reflected in the plausibility of countersocial counterfactuals for different people and situations. Some countergender reasoning is common and straightforward: 'If I had been a man, that person would not have cut in front of me in line' and 'If I had been a cis woman, that store would have treated me differently' are common sorts of thoughts. Yet many people encounter imaginative resistance when entertaining careful, deep countergender reflections: their lives might have been so different had they belonged to a different social category that they would no longer have been themselves. Whether or not such imaginative resistance tracks metaphysical facts is a question worthy of independent investigation. What is important for the present purposes is the idea that different people have different social categories essentially.

If the essentiality of social categories can differ from person to person, then a particular kind of countersocial, entertained by and about different persons, appeals to impossible worlds in some cases and merely possible worlds in others. I call this phenomenon *impossibility variance*. Impossibility variance is a feature of some properties, like gender, in which a property is essential to some, but not all, individuals.¹² If there are instances of impossibility variance, they

made them who they are if they had come from a different zygote. If Parfit's (1984: 387–389) origin essentialism is true, then 'If I had come from a different zygote...' is always a counterpossible.

¹¹ See Raven (2022) for relevant broader discussion about essences of social entities.

¹² For a detailed technical discussion of the differences between these sorts of cases, see Weinberger (2022). For a relevant law-focused discussion, see Dembroff, Kohler-Hausmann, and Sugarman (2020).

are a distinctive mark of countersocials, because counterlogicals and countermathematicals universally appeal to impossible worlds. ‘If two plus two had equaled five, small children wouldn’t have cared’ is a counterpossible no matter what owing to the impossible antecedent. A particular countersocial, however, might vary in its impossibility from case to case. A counterfactual that must appeal to impossible worlds in the case of one individual might only have to appeal to possible worlds in the case of another.¹³ ‘If I had been a straight woman, then I would not have left Kentucky’ is a counterpossible in the mouth of some individuals, but not in others.

As Wilhelm (manuscript) notes, many trans people take gender identification to be an essential part of who they are. But many other individuals do not, and historically, taking gender to be an essential property of individuals has contributed to marginalisation of women and trans people. The solution, according to Wilhelm, is to take gender to be an essential property of some, but not all, individuals. Some can switch genders and remain themselves; others would lose their essences and no longer be themselves. This view tracks the apparent variance in essentiality of social categories across individuals.

Accepting impossibility variance accommodates differential imaginative accessibility of countergender possibilities: for some people, entertaining a switch in their own gender is easy and imaginatively accessible. For others, the relevant imaginative activity runs contrary to essential identity properties. In the latter sort of case, entertaining a countergender counterfactual will be more like entertaining a countermathematical counterpossible, psychologically speaking: there will be limits on the coherence and detail of one’s imaginative projection. See also Cameron (2022: §4.3.3) for a discussion of some personality traits and personal relationships as essential to some individuals.

Gender is not the only social category with a complex relationship to metaphysical essentiality. Some disability theorists take particular disabilities to be essential to those who have them, especially given the social construction of aspects of disability that go along with physical

¹³ Some countersocials are even counteridenticals, which run contrary to identity facts. For example, ‘If I had been Frida Kahlo, I would have had more pet monkeys’ runs contrary to identity facts. So does ‘If Frida Kahlo had been Diego Rivera, she would have been taken more seriously as an artist in her own time’. First personal counteridenticals are also common in our reasoning: ‘If I had been my mother, I wouldn’t have changed my last name’ or ‘If I were you, I would withdraw the application’ have clear meanings and intuitive truth values. Even if these instances are read as non-counteridenticals that merely pick up on relevant aspects of similarity, they obey the same rules as many other counterpossibles. For an extended discussion of counteridenticals, see Kocurek (2018).

features. Others take certain disabilities to be essential to social cultures, like Deaf culture. Like gender, there are normative and theoretical reasons to accept impossibility variance such cases. Sinclair (2013) dislikes person-first language (for example, ‘person who is autistic’ or ‘person who is Deaf’) because he feels his identity cannot be separated from his autism. Sparrow (2005) views the pressure to adopt cochlear implants as an attempt to erase Deaf culture, including its signed and spoken languages. These claims can plausibly be interpreted as positing a sort of essentialism about people’s disabilities, and the natures of disabled communities. In the context of philosophy of religion, Timpe (2020) holds that some disabilities are essential to some individuals, and might accompany them to the afterlife. In some of these cases, people who are disabled are not separable from their disabilities. In other cases, people who are disabled are not separable from disabled cultures and communities to which they belong.

Another tricky aspect of countersocial evaluation is that membership in multiple social categories can be causally related. The sex columnist Dan Savage, for example, muses that he would not have given up his membership in the Catholic community if he had not been gay. (Savage 2013) In this sort of case, the nearby worlds in which one social category is altered are also worlds in which other, intertwined social categories are automatically altered. Intersectional social categories, which are explanatorily unified, sometimes require counterpossible evaluation when separating socially entwined categories. (In section 6, I introduce and discuss intersectional social categories in more detail.)

Evaluating countersocials is challenging in a number of ways. First: as mentioned above, distance from social actuality may be different than distance from physical actuality. While there is a relationship between physical and social actuality—that is, a change in physical reality will result in a change social reality—it is not clear that the standard Lewisian metric is effective at adjudicating the truth of countersocials. It is unclear, for example, how much of physical reality one must hold fixed in order to hold fixed corresponding aspects of social reality. Second: even if one is a friend of the truth and non-vacuity of counterpossibles, there is a natural question about how to adjudicate their truth: how does one evaluate distance from actuality when impossible worlds are in the mix? Here I will not attempt to provide a semantics for countersocial counterpossibles, though the area is ripe for future investigation. But I will make a few remarks about potential avenues of progress.

Ideally, the friend of countersocials should have a unified semantic treatment of non-

counterpossible and counterpossible countersocials. Countersocials with impossible and possible antecedents are treated similarly in everyday reasoning. This is an important difference between countersocials and other sorts of counterpossibles. Entertaining ‘If two plus two had equalled five, the laws of physics would have been different’ is quite different from entertaining ‘If Neil Armstrong hadn’t walked on the moon, the space program wouldn’t be as well-considered’. Whatever the context, the former requires a bit of intellectual and logical wrangling that the latter does not. In contrast, entertaining a countersocial like ‘If I had been a man, that meeting would have been easier’ usually doesn’t require or involve reflection about the rules of logic. In the ordinary course of thought, we do not pause and ask ourselves whether a countersocial scenario is possible or impossible.

A good unified semantic approach involves both hyperintensionality and impossible worlds. Hyperintensionality arises when there are necessary equivalents that are not intersubstitutable *salva veritate*. For example, someone might believe that two plus two equals four, and yet not believe that four equals four. But every possible world at which two plus two equals four is also a world in which four equals four: the two propositions are necessarily equivalent. A hyperintensional framework draws finer-grained distinctions than intensions.

There are many hyperintensional distinctions to be drawn in the social world. For example, a conspiracy theorist might believe that Barack Obama is Barack Obama, but not believe that Barack Obama was conceived by his parents Ann Dunham and Barack Obama Sr. Supposing that Obama’s conception by his parents is essential to him, every possible world that has Barack Obama includes his conception by Ann and Barack Sr.

Impossible worlds are a natural way to make sense of hyperintensional claims, since impossible worlds can model ways things could not have been. For origin essentialists, a Barack Obama who is not conceived by his particular parents only occurs at impossible worlds. And supposing that Donald Trump is essentially a man, only impossible worlds will include a Donald Trump who is not a man.¹⁴

Making sense of countersocials requires including impossible worlds in the set of worlds to be evaluated. Consider the claim ‘If Donald Trump had been a woman, he would have lost the

¹⁴ Note that one need not believe in impossible worlds in order to be onboard with non-trivial countersocial evaluation. Even if impossibilities turn out to be inconsistent sets of propositions or other sorts of models, countersocial evaluation sometimes requires appeal to the impossible.

2016 election'. Making sense of this claim requires appealing to the impossible world where Trump is a woman. Like ordinary counterfactuals, we can appeal to contextually defined standards of similarity between worlds in order to evaluate the claim. In this case, we hold many of the other relevant social facts fixed (including the particular socio-political situation of the United States in 2016, Hillary Clinton as Trump's opponent, the tendency to judge women candidates harsher than candidates who are men, etc.) and just vary Trump's gender. And it is plausible that Trump would not have won nor even have been nominated if he were a woman (Shayne 2020). While there might be problems with evaluating countersocials within the traditional similarity metric, including impossible worlds is a step in the right direction.

The friend of counterpossible countersocials should be comfortable violating Nolan's (1997: 536–40) Strangeness of Impossibility Condition (SIC). Roughly, SIC holds that any possible world is closer to the actual world than any impossible one. The intuition behind SIC is that any impossible world will be stranger and more different than the actual world than any possible one. But many counterpossible countersocials seem to draw on dimensions of similarity that intuitively order impossible worlds closer to the actual world than possible worlds. For example, worlds in which Donald Trump is a woman (and lost the election) seem closer than worlds in which Prince William won the 2016 presidential election.

A hyperintensional framework with impossible worlds also permits neutrality on the question of whether or not membership in certain social categories is an essential feature of individuals. Consider a person, Jacinda, who entertains the idea that had she not been a woman, she might have studied engineering. Jacinda herself might not know if her gender identity is essential to her. A third-personal evaluation might yield no more modal knowledge in the matter.¹⁵ For many countersocial evaluations, it is unclear whether one is entertaining a counteressential or not. Restricting the evaluation to possible worlds forces a resolution on the essentiality of social traits, since counteressentials appeal to more than possible worlds. A unified framework for possible and impossible countersocials accommodates this natural uncertainty in countersocial reasoning.

Finally, a framework with impossible worlds accommodates intersectional social categories, which are characterised by multiple interacting and inseparable axes of social

¹⁵ For the purposes of this paper, I remain neutral on theories of personal identity; in particular, I remain neutral on the merits of narrative theories of identity versus other sorts of theories.

oppression. *Black womanhood* is Kimberlé Crenshaw's (1989: 149–52) canonical example of intersectionality: Black women face types of sociopolitical oppression that are greater and different than the sorts of oppression faced by Black people and by women. The interactions of the types of oppression yield a form of oppression distinctive of the intersectional social category. Intersectionality can also be construed as a claim about the dependent and intertwined existences of some social categories: in some cases there cannot be a particular gendered social category without its being racialised, and vice versa.

There is some debate about how to model the metaphysics of intersectionality (Garry 2011; Jorba and Rodó-Zárate 2019). Here I follow Bernstein (2020: 330–334) in holding that intersectional social categories are explanatorily and metaphysically unified, and downwardly ground their constituent identity categories. Hyperintensionality is a useful tool in accounting for unified intersectional social categories that might otherwise be inseparable in a merely intensional framework. For example, suppose that a Black woman is trying to assess whether or not she would have been called 'angry' by her employer if she was not a woman. And suppose that the employer has a particular animus towards Black women (but not towards Black men, and not towards non-Black women.) If *Black womanhood* is a metaphysically unified social category in this instance, then one cannot separate Blackness from womanhood within the intersectional category in merely possible worlds. Any world in which the employer discriminated against people who are Black is also a world in which the employer discriminated against people who are women, since the discrimination involves the unified category *Black woman*. Impossible worlds are useful in evaluating whether the label would have been applied in the absence of Blackness or womanhood, in contexts where the two categories are in some sense inseparable.¹⁶ I discuss intersectional oppression in more detail in section 6 below.

Taking stock of the terrain: I have suggested that some countersocial counterpossibles can be true and non-vacuous. A hyperintensional framework for countersocial counterpossibles best handles impossibility variance, the idea that particular social properties can be essential to some, but not all, individuals. I turn now to connections between countersocials and the metaphysics of causation.

¹⁶ I assume here that it is metaphysically necessary *de re* of some Black women that they cannot be white women or Black men.

4. Causal Countersocials

It is widely thought that there is a strong connection between counterfactual dependence and causal dependence. While the exact nature of the relationship is controversial, counterfactual dependence does seem to capture an intuitive feature of causation: changing or removing a cause often changes or removes the effect (Lewis 1973a; 2000). For example, ‘If Takashi had not thrown the rock, the window would not have shattered’ captures an important feature of the relationship between the rock-throw and the shattering of the window: remove the former, and the latter does not occur. The causal relationship also captures covariance between cause and effect: altering the manner of occurrence of *c* alters the manner of occurrence of *e*. For example, if Takashi were to throw her rock one second earlier, the window should shatter one second earlier.

Drawing on this connection, I hold that the truth and non-vacuity of countersocial counterfactuals is *prima facie* evidence for a type of causal dependence between social categories and other events. True countersocials generate corresponding true causal counterfactuals like ‘If Janelle hadn’t been butch, she wouldn’t have been mistaken for a man’ and ‘If Manolo hadn’t been Hispanic, he wouldn’t have been hassled by security’. In such cases, removing or changing the social categories changes the relevant outcomes: If Janelle had been femme, she wouldn’t have been mistaken for being a man; and if Manolo had been white, he would not have been singled out by security.¹⁷

I suggest that social categories can also be causal *relata*: social categories like *Black* and *woman* are literally causal. Traditional metaphysical theories of causation focus on specific token events as the causal *relata*; for example, driving through a red light causes a traffic accident (Lewis 1973a; 2000). But if the idea of causation as a type of intervention is to be taken seriously, it is natural to hold that counterfactual variation of events with social categories tracks some sort of causation.¹⁸ Social categories centrally feature in predictive and explanatory claims, and social categories produce true corresponding causal counterfactuals. They do so in two ways. First, membership in social categories is straightforwardly causal. Consider these causal claims,

¹⁷ See Hu (forthcoming) for a broad argument for the interdependence of causation with normative features, including race and gender. See also Faller (2023) and Silver (2024) for other examinations of social causation.

¹⁸ Here I remain neutral on which account of counterfactual causation is the correct one, which does not matter for these purposes.

paired with corresponding countersocials:

Being a woman causes Dahlia to receive 20% less pay for the same job than her male colleagues.

* If Dahlia hadn't been a woman, she would have received 20% more pay.

Being a Black man causes Jamal to be pulled over by police.

* If Jamal hadn't been Black, he wouldn't have been pulled over by police.

Shay's being a trans man causes people to awkwardly use 'Yasss! King' as an affirmation for Shay.

* If Shay hadn't been a trans man, people wouldn't awkwardly use 'Yasss! King' as an affirmation.

Sam's being obviously American affected his ability to gain entrance to Berghain.¹⁹

* If Sam hadn't been obviously American, he would have gained entrance to Berghain.

Second, social categories themselves— even if they are abstracta— are causal. For example:

Changes to the social category of marriage have meant greater rights for LGBTQ+ individuals.

* If changes to the social category of marriage had not occurred, LGBTQ+ would lack certain rights they now have.

Social categories can be effects as well as causes. For example:

¹⁹ Berghain is a Berlin nightclub famous for its stringent entrance requirements. The requirements are interpreted and imposed by a notoriously picky bouncer.

The passage of the Johnson-Reed Immigration Act of 1924 caused the category of ‘Asian’ to become more monolithic.

* If the Johnson-Reed Immigration Act of 1924 had not passed, the category of ‘Asian’ would have been different.

Audre Lorde and Kimberlé Crenshaw redefined Black womanhood.

* If Audre Lorde and Kimberlé Crenshaw had not written about Black womanhood, the category would not have been legally codified.

Social categories can also be causal intermediaries. For example:

Callie and Jacintha couldn’t get married in 1990, which caused them to join the effort to expand the category of marriage to same-sex marriage. Laws allowing same-sex marriage were passed in 2009, prompting them to get married in 2010.

In each of these cases, altering the social category generates counterfactual variance in the outcome a pattern from which causal implications can be drawn.²⁰

Social categories can be preemptors, overdeterminers, and joint causes.²¹ Roughly, there is a case of *preemption* when there are multiple sufficient causes for an outcome, and one cause prevents the other from bringing about the effect. There is a case of *overdetermination* when there are multiple independent sufficient causes for an outcome. And there is a case of *joint causation* when multiple causes are jointly necessary to bring about an outcome. Consider:

²⁰ I take such dependence to be causation rather than some other sort of ontological dependence like grounding, supervenience, material constitution, parthood, or the like, because the cases with which I am concerned are best explained by causal relationships rather than other sorts of relationships. For example, the relationship between being Black and being pulled over by police is not best explained by material constitution or by parthood, though such relationships might be involved in the existence of social facts more generally. And following Bernstein (2016) and Koslicki (2016), I hold that causation is different from grounding.

²¹ This approach is friendly to the interventionist causal program, which I take to be a similar and compatible approach to these topics.

Social Preemption: Aroha, a disabled Maori woman, applies for a work visa in Australia in 1920 and is denied. If she had not been denied the visa because she was Maori, she would have been denied the visa because she was disabled.

Social Overdetermination: Dvora, a Jewish and Republican-voting woman, messages a romantic prospect on Tinder. The prospect dislikes both Republican-voting women and Jewish women: either property would have been sufficient for rejection. The prospect ignores Dvora's message.

Social Joint Causation: Amit, a Sikh doctor with no financial credit, applies to a bank for a mortgage loan that requires the approval of two loan officers. One loan officer has benevolent bias toward people who are Sikh, and the other loan officer has benevolent bias towards medical doctors, and the two officers extend the loan.

In each case, social categories play the explanatory and predictive roles that events normally do. In the preemptive and jointly caused cases, altering the social categories alters the outcomes. And in the overdetermination case, each social category is independently sufficient to predict and explain the outcome.

The metaphysical claim that social categories are causal aligns with the way that social categories are utilised in other academic fields. Relevant social sciences already make heavy use of social categories as causal entities. It is widely accepted in economics, for example, that varying race and gender variables will vary economic outcomes. (Greenman and Xie 2008) In public health, membership in marginalised racial identities in the United States are the top predictors of adverse outcomes for diabetes. (Walker, Williams, *et al.* 2016) And psychologists agree that children are racialised and gendered in ways that affect their subsequent coping mechanisms and mental health outcomes. (Pachter and Cole 2009)

Understandably, there is resistance to the idea that social categories themselves are causes. The thinking is that social categories are socially constructed: they don't seem to have the same physical *oomph* as, say, falling dominoes. The existence and natures of social categories are dependent on human thoughts and activities. And social categories do not carve reality at its physical joints.

Thus one might be tempted to hold a different, weaker theoretical alternative. One might hold that it is *perceptions* of membership in social categories, rather than social categories themselves, that do the causal work. For example, when Manolo is hassled by security, it is because he is perceived to be Hispanic. And when Janine would have had an easier time in a meeting if she were not a woman, the trouble is due to her being perceived to be a woman. This sort of view locates causal power of social categories in perceptions of group membership, rather than social categories themselves.

This temptation should be resisted for several reasons. First, entities need not be mind-independent in order to be causally efficacious. Socially constructed entities, like money and art, are often taken to be straightforwardly causal. Spending money, for example, causes one to receive goods. A painting's classification as art causes it to have certain properties that a mere painted canvas does not. Like social categories, art and money generate uncontroversially true counterfactual causal claims. Socially constructed entities are part of the causal fabric of the world.

Second, claims about perceptions of social categories are not always intersubstitutable *salva veritate* with claims about social categories themselves. Consider 'If Cane hadn't been perceived by their friends as non-binary, they wouldn't have struggled with what sort of tie to wear' versus 'If Cane hadn't been non-binary, they wouldn't have struggled with what sort of tie to wear'. Supposing that Cane's indecision about their tie is private to them, external social perception does not play a role in their deliberations.

Another reason is that even in cases where the perceptual claim comes out true, it doesn't give us the whole explanation. Suppose that Sunita, a Dalit woman in India, is sad about her caste-restricted marriage prospects. Sunita would not bemoan the fact that she is *perceived* as Dalit, though that obviously plays a causal role in her social situation. Being perceived as Dalit isn't the best explanation of Sunita's limited nuptial prospects: it is, rather, membership in the category itself, including having been born into the category.²² Being Dalit explains why she is perceived as Dalit. Further, the social category itself backs explanations better than relevant alternatives, like reductions to lower-level physical causal processes.

²² See Bernstein (2024b: 86-87) for more discussion of this sort of example, including the restrictions on agency imposed by the social categories into which one is born.

An additional alternative is to locate the causal power of the social world in the possession of social properties rather than membership in social categories. For example, possessing the social property *Black* causes Jamal to be pulled over, rather than the social category *Black* doing the causal work; and having the property *Dalit woman* influences Sunita's marriage prospects, rather than the social category itself being the source of the causal power.

This approach is not substantively different from the one that I propose. What is important for my discussion is that social categories play central causal roles in the lives of their members, however the precise ontological details of the categories are hashed out. While I take these details to be of metaphysical interest more broadly, I will not further argue for the causal power of social categories here. Even those who resist the claim that social categories are causal can be open to the idea that social categories participate in intuitively true causal counterfactuals, and that social categories back predictions, explanations, and causal claims.

5. Causal Countersocials and Impossible Worlds

Following Nolan (1997), Brogaard and Salerno (2013), Bernstein (2016), Baron and Colyvan (2021), Tan (2019), and Berto and Jago (2023), I hold that impossible worlds can be used in causal explanation. Given the close relationship between explanation and counterfactual dependence, impossible worlds are involved in explanations that appeal to impossibilities. For example, in the metaphysics of mind, there is general agreement among physicalists that a physical property, *P*, cannot occur without its supervening mental property, *M*. In this sort of case, the most informative counterfactuals are ones that appeal to impossible worlds, like 'If *P* occurred without *M*, Jane's hand would not have been raised'. The impossible world, the one where *P* occurs without *M*, is more informative than the possible worlds in which *P* and *M* co-occur. Tan (2019) provides another example of the explanatory indispensability of impossible worlds within physical explanations. The counterpossible 'If diamond had not been covalently bonded, it would have been a better electrical conductor' is key to explaining the conductivity of diamond, since diamond is necessarily covalently bonded. An impossible world without a covalently bonded diamond provides an informative contrast with the actual world. Baron and Colyvan (2021: 6) implicate the explanatory power of impossible worlds in physics, citing the truth of 'If the speed of light had been 302,000 km/s, it would have been possible for an object

with positive rest mass to accelerate to 300,000 km/s'. *Mutatis mutandis* for other topics in which counterpossibles are relevant, including metaethics (Bernstein 2016: 2586–87), philosophy of science (Wilson 2021), and the philosophy of computability (Jenny 2018).

Similarly, impossible worlds play several key roles in the explanatory power of countersocials. As discussed above, impossible worlds make sense of situations in which identity facts or essentiality facts are violated. For example, 'If Donald Trump had not been a man, the campaign would have been run differently' appeals to counteressential worlds in which Donald Trump is not a man. Impossible worlds are additionally informative via contrasts between possibility and impossibility. Contrasting the above counteressential with 'If Donald Trump had been born in California, the 2016 campaign would have been run differently' is informative insofar as the latter does not need to appeal to impossible worlds. Arguably, impossible worlds can also help make sense of apparent true contradictions in the social world. (Cull and Bolton 2019)

Impossibilities also play a role in causation. Following Bernstein (2016), I hold that there can be impossible omissive causes—causally relevant events that didn't happen, but could not have happened in any case. Suppose that Jamina, a tenure-track mathematics professor, attempts to prove a mathematical impossibility (say, that two plus two equals five), and fails. And further suppose that if she hadn't failed to prove that two plus two equals five, she would have received tenure. Such impossible omissive causes are reflected in the truth of causal counterpossibles. For example, 'If Jamina hadn't failed to prove that two plus two equals five, she would have received tenure' is true partly because something occurs at an impossible world—Jamina's proof that two plus two equals five. In cases of metaphysically impossible omissions, in which there are events that did not happen but could not have happened, omissive causes can only be modelled by making use of impossible worlds. The causal claim is made true partly by worlds in which the impossible event (in this case, Jamina's proof that two plus two equals five) occurs.

Counterpossible countersocials can correspond to true omissive causal claims. Consider 'If Gandhi had not lacked a particular social status, he would not have become an activist'. Assuming that Gandhi essentially had a particular social status, his failure to belong to a higher status social category caused him to become an activist. Or consider 'If Helen Keller had not lacked the ability to see and hear, she would have had an entirely different life'. If we assume

that Helen Keller had her disabilities essentially, one must appeal to impossible worlds in order to evaluate the countersocial claim.

In addition to omissive causal claims, countersocial counterpossibles also generate true contrastive causal claims. Contrastive causal claims take the form ‘*c* rather than *c** caused *e* rather than *e**’, where *c* and *e* are actual events and *c** and *e** are non-actual contrasts. (Schaffer 2005) ‘The milk (rather than the bread) caused the allergic reaction (rather than the healthy digestion)’ is a typical contrastive causal claim. Counterpossible countersocials can generate such a claim appealing to impossible contrasts to the cause, the effect, or both. For example: suppose that Juan, a Latino man, could not have been a white woman. And suppose that Juan is hassled by police because he is a Latino man. Then the following contrastive claim is true:

Juan’s being a Latino man (rather than a white woman) caused him to be hassled by the police (rather than not be hassled).

Evaluating the truth of the claim requires appeal to an impossible contrast—one in which Juan is a white woman. Not all contrasts are impossible. Juan might have been a Mestizo man, for example, even if he could not have been a white woman. Impossibility variance in contrastive causal claims is another reason to accept a hyperintensional framework for countersocials.

More generally, causal counterpossibles can be non-vacuously true or false when some essential social feature is doing causal work. For example, suppose that Milton, who is essentially a white man, receives a job offer because he is a white man. The causal countersocial ‘Were Milton not a white man, he would not have received the job offer’ cannot be evaluated without appeal to impossible worlds, since Milton is a white man at every possible world. As with other sorts of counterpossibles, there are intuitive differences in truth values between countersocials. For example, ‘Were Milton not a white man, plants would not have sprouted’ seems false.

Impossible worlds are also helpful in making sense of grounding claims in social metaphysics, as many philosophers hold that social construction is a form of grounding (Epstein 2015; Griffith 2018a and 2018b). *Grounding*, the non-causal relationship between more and less fundamental parts of reality, is widely thought to be a metaphysically necessary relationship: if *x* fully grounds *y*, *x* grounds *y* necessarily (Schaffer 2016; Bernstein 2016). For example, if an

arrangement of particles fully grounds the existence of a particular baseball, this grounding relationship is present in every possible world where the arrangement of particles exists (Trogon 2013). While for the physicalist it is obvious that the social world is somehow grounded in ‘particles [and] fields’ (Schaffer 2017: 13), there is disagreement over the intermediate steps in the grounding chain between the microphysical world and the social world—specifically, which social patterns and institutions give rise to categories like gender and race.²³

Due to the necessity of the grounding relationship, claims about grounding often utilise counterpossibles (Wilson 2018). When inquiring as to whether a group of particles ground the existence of a particular baseball, for example, one would appeal to the impossible world in which there are particles but no baseball. Similarly, debates over the grounds of gender must also appeal to impossible worlds in which there are social categories like gender, but in which there are no grounds or different grounds than the actual world. These inquiries involve countersocial counterpossibles because they entertain scenarios that violate metaphysical necessity.

6. Intersectionality and Causation

Finally, we can draw together the threads of the preceding discussion in ways that are useful for understanding intersectional oppression. Suppose that social categories are causal in the way that I have suggested. There is an additional question about how intersectional social categories like *Black womanhood* cause other things.

Yablo (1992) has the following useful take on causation: for c to cause e , c must be proportional to e . Roughly, c is proportionate to e if c has the appropriate level of causal detail (Sartorio 2010; Dowe 2010; Bernstein 2014; McDonnell 2018; Touborg 2022; Rubenstein 2024a and 2024b). Levels of causal detail are differentiated by the determinate/ determinable relationship, a hierarchical relationship of ascending specificity that holds between properties like *fuchsia* and *pink*, and *square* and *shaped*. For example: suppose that Sophie the bird is trained to peck only at pink objects, and Sophie pecks at an object that is fuchsia. The candidate causes for Sophie’s pecking are:

²³ Schaffer (2017: 6), for example, holds that ‘To be socially constructed is to be grounded in distinctive social patterns.’

- (1) *the object's being coloured*
- (2) *the object's being pink*
- (3) *the object's being fuchsia*

Fuchsia is a determinate of the determinable *pink*, and *pink* is a determinate of the determinable *coloured*. Causal claims are to be counterfactually tested to evaluate whether they involve the correct level of detail. For example: is it the case that had the object not been fuchsia, Sophie would not have pecked? No, because Sophie would have pecked even if the object had been bubblegum pink. Is it the case that had the object not had any colour, Sophie would not have pecked? No, because the object needs to have a colour—pink—in order to be pecked. The cause of the pecking is (2): *the object's being pink*, rather than the object's being fuchsia or coloured, because the object's being pink is the correct level of causal detail. The determinable *being pink* is the cause of Sophie's pecking. In general, proportionate causes can't be too specific for their effects ('Milo was run over by a polka-dotted yellow school bus!') or too broad for their effects ('Milo was run over by an object extended in spacetime!')

Consider another example of proportionate causation. Suppose that the only supermarket that makes Dvora happy is the Trader Joe's on Hyde Street in San Francisco, California. And suppose that Dvora goes shopping at that very Trader Joe's. The candidate causes for Dvora's happiness are:

- (1) *being a supermarket*
- (2) *being a Trader Joe's*
- (3) *being a Trader Joe's in San Francisco, California*
- (4) *being a Trader Joe's on Hyde Street in San Francisco, California*
- (5) *being a Trader Joe's on Hyde Street in San Francisco, California on a sunny day*

As before, each claim involving a determinate can be counterfactually tested. Is it the case that going to any supermarket, or any Trader Joe's, or any Trader Joe's in San Francisco, would have made Dvora happy? No: only the Trader Joe's on Hyde Street would have made her happy. Nor

is it a happiness requirement for Dvora that it is a sunny day at the Trader Joe's on Hyde Street. The cause of Dvora's happiness is (4), since it contains the correct level of causal detail.

With this framework in mind, reconsider the concept of intersectionality. Originating with Kimberlé Crenshaw (1989) and the Combahee River Collective (1977), the general idea of intersectionality is that different axes of social oppression intersect and interact, creating distinctive new forms of oppression. According to Crenshaw, 'the intersectional experience is greater than the sum of racism and sexism.' (1989: 140) In more of her words:

Many of the experiences Black women face are not subsumed within the traditional boundaries of race or gender discrimination as these boundaries are currently understood, and [the] intersection of racism and sexism factors into Black women's lives in ways that cannot be captured wholly by looking at the women, race or gender dimensions of those experiences separately. (Crenshaw 2006: 8)

Intersectional oppression is *non-additive* in that it is different from the mere joint occurrence of racist and sexist oppression. Intersectionality describes cases in which members of multiple marginalised social groups face forms of oppression based specifically on the interaction of the types of marginalisation.²⁴

Discrimination is a common form of intersectional oppression. In Crenshaw's famous *Moore versus Hughes Helicopter* example, a Black woman sued Hughes Helicopter Corporation for discrimination on the grounds that Black women were not being promoted at the company. The court ruled that the discrimination did not occur on sexist grounds (since women had been promoted) and did not occur on racist grounds (since Black men had been promoted). As Crenshaw famously pointed out, the category *Black woman* was not recognised by the court as a protected class because the court failed to take into account the intersection of racist and sexist oppression.²⁵

²⁴ According to the Combahee River Collective, 'the major systems of oppression are interlocking. The synthesis of these oppressions creates the conditions of [Black women's] lives.' (1977: 1; Taylor 2017 edition)

²⁵ For ways that social properties ground the existence of certain social categories but not others, see Mason and Ritchie (2020). For an elaboration and discussion of 'ontic injustice', a sort of injustice resulting from social categorization, see Jenkins (2020).

I suggest that proportionate causation is a helpful framework for modelling intersectional oppression. Consider Crenshaw's *Moore versus Hughes Helicopter* example. Differentiating by determinables and their determinates, some candidate causes for discrimination are:

(1a) *being Black*

(2a) *being a Black woman*

(1b) *being a woman*

(2b) *being a Black woman*

As in our initial cases of proportionate causation, we can ask: is it the case that if the plaintiff had not been Black (but had been a woman), she would not have been the target of this kind of discrimination? And is it the case that if she had not been a woman (but had been Black), she would not have been the target of this kind of discrimination? The answer to each is no, for if she had been either a man or a non-Black woman, she likely would not have faced this sort of discrimination. *Being a Black woman* is the cause of the intersectional oppression, because it occurs at the appropriate level of causal detail.²⁶ The counterfactual dependence obtains between the category *Black woman* and the discriminatory behavior. Proportionality is used to select among multiple candidate counterfactual claims of social dependence.²⁷ Whether or not a candidate cause is the proportionate one is best evaluated in terms of counterfactuals, a feature which establishes the link between social dependence and proportionality.

As an additional example, consider a Muslim woman, Zainab, who is sniped at for wearing a hijab at work. Appealing separately to Zainab's womanhood or Zainab's being Muslim does not capture the phenomenon: hijab-wearing Muslim women are stereotyped in particular ways that Muslims (in general) and women (in general) are not. The intersection of the social categories is the locus of Zainab's oppression. It is the unified social category—*Muslim woman*—that explains Zainab's particular form of oppression. As above, we can use

²⁶ Kofi Bright, Malinsky, and Thompson (2016) also offer a causal theory of intersectionality, as represented in causal models. I take my point about proportionality to be complementary to this and similar views.

²⁷ This connection between proportionality and counterfactuals reflects the approach of List and Menzies (2009), who hold that a cause is proportional only if it covaries with its effect in contextually relevant situations. See also Rubenstein (2024a) and (2024b) for formulations of the link between counterfactuals and proportionality.

proportionality as a preliminary test for intersectional oppression. Candidate causes for the sniping are:

- (1a) *Zainab's being a woman*
- (1b) *Zainab's being a Muslim woman*

- (2a) *Zainab's being Muslim*
- (2b) *Zainab's being a Muslim woman*

Being neither a woman nor being Muslim is the correct level of detail for the causal claim, since neither men nor non-Muslims wear a hijab. *Being a Muslim woman* is the appropriate level of detail to describe the intersectional oppression.

Proportionate causation can also provide a useful test for whether the oppression in question is based on membership in an intersectional or unitary social category. In many cases, discrimination is based on only one social category. Consider an example in which Caily, a lesbian wheelchair user, is often excluded from meetings because their building's elevator is broken. What is the cause of their exclusion? Differentiating by determinables and their determinates,²⁸ some candidate causes of discrimination are:

- (1a) *Caily's being a wheelchair user*
- (1b) *Caily's being a lesbian wheelchair user*

- (2a) *Caily's being a lesbian*
- (2b) *Caily's being a lesbian wheelchair user*

In this case, (1a) is the cause of the exclusion, since it has the appropriate level of detail. Being a lesbian is presumably irrelevant to the exclusion. Caily's oppression in this case is not intersectional because it involves a unitary social category.

²⁸ Note that one need not adopt the framework of determinates and determinables in order to utilise proportionate causation, though I do so in order to adopt Yablo's classic theory of proportionate causation.

Proportionate causation is not a sufficient condition for intersectional oppression, however. Intersectionality involves not just numerous co-instantiated social categories, but also by definition involves their *intersection*-- some sort of significant interaction, mutual constitution, or tight relationship between the categories. Intersectional oppression is often characterized as non-additive: adding oppression based on Blackness to oppression based on womanhood, for example, does not fully add up to intersectional oppression of Black women, because it does not take into account the ‘mixing’ aspect of this sort of oppression.²⁹ Describing intersectionality through metaphor, Haslanger (2013: 5) says ‘the intersection of race and gender has an effect similar to overlapping different colored gels on a theater light.’

To see why proportionate causation is not the whole story, consider the following case:

Building Entrance

Jacquise is a Black wheelchair user. He wants to enter a building with two doors. The front door is disability-accessible, and the side door is not. Jacquise is denied entry by the security guard posted at the front door, who is suspicious of Black men. Jacquise cannot enter through the side door, since it is not disability-accessible.

In this case, *being a Black wheelchair user* is the proportionate cause of Jacquise’s failure to enter the building. But the social category *Black wheelchair user* is not a locus of intersectional oppression: the two social categories do not cooperate to produce oppression in this case. Jacquise’s oppression in this case just is a result of two co-instantiated social categories.³⁰

Consider another example of proportionate causation without intersectional discrimination:

²⁹ There is a separate question about the metaphysical and explanatory relationship between the individual social categories in these cases. See Bernstein (2020), Jorba and Rodó-Zárate (2019), and Dembroff (2024) for general metaphysical theories of intersectionality.

³⁰ There are similar instances in which these very same co-instantiated social categories lead to intersectional oppression. For example, in a variant of **Building Entrance** in which the security guard is ill-disposed to Black wheelchair users because of deeply held stereotypes about them, this might count as a case of intersectional oppression. The requirement for intersectional oppression includes interaction of the social categories in addition to co-instantiation. The initial **Building Entrance** case is characterised by mere co-instantiation of the social categories.

Startup

A startup company tends to hire only Millennials with Ivy League degrees into its junior engineering positions. Ricardo, a highly qualified 55-year-old baby boomer with a degree from Kansas State University, applies to a position at the startup and does not get the job.

In this case, Ricardo faces discrimination based on both his age and his educational attainment, but does not face discrimination based on their intersection. This is an example of discrimination involving multiple social categories that is not intersectional. It is, rather, a case of social overdetermination: both Ricardo's age and his educational attainment are individually sufficient to prevent him from being offered the job. But the two categories do not interact in a significant way.

This type of contrast case, which we can term a mere co-instantiation case, helps model what is distinctive about intersectional oppression. When multiple social categories give rise to a sort of discrimination that occurs due to an interaction of categories, like *Black womanhood*, it is an instance of intersectional oppression. Where multiple co-instantiated categories separately cause discrimination, but without significant interaction³¹ between the co-instantiated categories, it is not a case of intersectional oppression. The fact that proportionate causation is not sufficient for intersectional oppression additionally illuminates the notion of intersectionality: the relationship between social categories plays a central explanatory and metaphysical role.

The key question is what sort of relationship between social categories is required for intersectional oppression in particular cases. Though the details of a positive account would require extended discussion, there are numerous accounts on offer. Bernstein (2020) suggests that the relationship between intersectional categories is one of explanatory and metaphysical unity. Jorba and Rodó-Zárate (2019) describe intersectional categories as emergent experiential entities. Garry (2011) describes intersectionality through the metaphor of mixed liquids.³² Dembroff (2024) characterises intersectionality through overlapping systems of oppression. Even

³¹ Here I do not intend to take a stand on whether so-called interactionist models of intersectionality are correct. I use 'interaction' to broadly cover a variety of possible relationships between social categories.

³² '[Liquids] show the ways in which some oppressions or privileges seem to blend or fuse with others. Different liquids—milk, coffee, nail polish, olive oil, beet borscht, paint in several colors—run down from different places at different altitudes into roundabouts. Some of the liquids run together, some are marbled with others, and some stay more separate unless whipped together.' (Garry 2011: 405)

empirical data can help determine when intersectional oppression occurs (Bauer, Churchill *et al.* 2021) Each approach is compatible with the idea that proportionate causation is a central part of the phenomenon, but not the entire phenomenon. Further investigation of the topic would be fruitful. What I have shown is that we need countersocial counterfactuals to understand proportionality in a social context, and we need proportionality to grasp what is still missing in the metaphysical investigation of intersectionality.

7. Concluding Remarks

Countersocial counterfactuals are an understudied phenomenon in the literature on counterfactuals and conditionals. Many countersocial counterfactuals are true, non-vacuous counterpossibles that we use in everyday reasoning about our lives. If we take the relationship between true counterfactuals and causal dependence seriously, we should hold that social categories are causal: they can be causes, effects, and causal intermediaries. These results bear on important topics in the philosophy of gender and race, including the apparent essentiality of social categories to certain individuals, and the metaphysics of marginalised identities. The truth and non-vacuity of countersocial counterpossibles suggests that impossible worlds are helpful in making sense of some instances of causation by social categories, as well as social construction as a form of grounding. Additionally, intersectional oppression can be partly illuminated by appeal to the proportionality of causation. Connections between social metaphysics and the metaphysics of causation are ripe for future development.

Acknowledgements

Thanks to Matthew Andler, Nina Emery, Vera Hoffmann-Kolss, Marta Jorba, Arc Kocurek, Jessica Li, Daniel Nolan, Dee Payton, Michael Rea, Kevin Richardson, Christopher Shields, Isaac Wilhelm, and audiences at UC Davis, UC Santa Cruz, UT Austin, University of Lisbon, Pompeu Fabra University, University of Arizona, Duke University, Mount Holyoke College, the University of Iowa, the University of Manchester, the University of St Andrews, UNC Chapel Hill, North Carolina State University, Umeå University, the University of Cologne, the 2023 meeting of the Australasian Association of Philosophy, and the 2023 meeting of the Central

Division of the American Philosophical Association in Denver for helpful feedback on this paper. I also thank several anonymous referees whose feedback improved this paper.

Funding Information

This project was supported by a 2022-2023 fellowship from the National Endowment for the Humanities.

ORCID

Sara Bernstein: ORCID 0000-0003-4288-9621.

References

Alderman, Naomi (2017) *The Power*. Little, Brown and Company.

Appiah, Anthony (1990) "But Would That Still Be Me?" Notes on Gender, "Race," Ethnicity, as Sources of "Identity", *Journal of Philosophy* **87**: 493–99. doi:10.5840/jphil1990871026

Ásta (2018) *Categories We Live By: The Construction of Sex, Gender, Race, and Other Social Categories*. Oxford University Press.

Baron, Sam and Mark Colyvan (2021) 'Explanation Impossible', *Philosophical Studies* **178**: 559–76. doi:10.1007/s11098-020-01445-8

Bauer, Greta, Siobhan Churchill, Mayuri Mahendran, Chantel Walwyn, Daniel Lizotte et al. (2021) 'Intersectionality in Quantitative Research: A Systematic Review of Its Emergence and Applications of Theory and Methods', *SSM – Population Health* **14**. doi:10.1016/j.ssmph.2021.100798

Bernstein, Sara (2014) 'Two Problems for Proportionality about Omissions', *Dialectica* **68**: 429–41. doi:10.1111/1746-8361.12071

Bernstein, Sara (2016a) 'Omission Impossible', *Philosophical Studies* **173**: 2575–89. doi:10.1007/s11098-016-0672-9

Bernstein, Sara (2016b) 'Grounding Is Not Causation', *Philosophical Perspectives* **30**: 21–38. doi:10.1111/phpe.12074

Bernstein, Sara (2020) 'The Metaphysics of Intersectionality', *Philosophical Studies* **177**: 321–35. doi:10.1007/s11098-019-01394-x

Bernstein, Sara (2024a) 'Biased Evaluative Descriptions', *Journal of the American Philosophical Association* **10**: 295–312. doi:10.1017/apa.2023.5

Bernstein, Sara (2024b) ‘Resisting Social Categories’, in David Shoemaker, Santiago Amaya, and Manuel Vargas, eds., *Oxford Studies in Agency and Responsibility*, vol. **8**: 81–102. Oxford University Press.

Berto, Francesco, Ellie Ripley, Graham Priest, and Rohan French (2018) ‘Williamson on Counterpossibles’, *Journal of Philosophical Logic* **47**: 693–713. doi:10.1007/s10992-017-9446-x

Berto, Francesco and Mark Jago (2023) ‘Impossible Worlds’, in Edward N Zalta and Uri Nodelman, eds., *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2023/entries/impossible-worlds/>

Bright, Liam Kofi, Daniel Malinsky, and Morgan Thompson (2016) ‘Causally Interpreting Intersectionality Theory’, *Philosophy of Science* **83**: 60–81. doi:10.1086/684173

Brogaard, Berit and Joe Salerno (2013) ‘Remarks on Counterpossibles’, *Synthese* **190**: 639–60. doi:10.1007/s11229-012-0196-6

Butler, Octavia (1993) *Parable of the Sower*. Four Walls Eight Windows.

Cameron, Ross P (2022) *Chains of Being*. Oxford University Press.

Colby, Tanner (2005) *Some of My Best Friends Are Black: The Strange Story of Integration in America*. Farrar, Straus and Giroux.

The Combahee River Collective, ‘The Combahee River Collective Statement’ (April 1977).

Cull, Matthew J and Emma Bolton (2019) ‘Contradiction Club: Dialetheism and the Social World’, *Journal of Social Ontology* **5**: 169–80. doi:10.1515/jso-2019-0035

Crenshaw, Kimberlé Williams (1989) ‘Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics’, *University of Chicago Legal Forum*: 139–67.

Crenshaw, Kimberlé Williams (1991) ‘Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color’, *Stanford Law Review* **43**: 1241–99. doi:10.2307/1229039

Dembroff, Robin (2024) ‘Intersection Is Not Identity, or How to Distinguish Overlapping Systems of Injustice’, in Ruth Chang and Amia Srinivasan, eds., *Conversations in Philosophy, Law, and Politics*: 383–98. Oxford University Press.

Dembroff, Robin, Issa Kohler-Hausmann, and Elise Sugarman (2020) ‘What Taylor Swift and Beyoncé Teach Us About Sex and Causes’, *University of Pennsylvania Law Review* **169**: 1–12.

Dowe, Phil (2010) ‘Proportionality and Omissions’, *Analysis* **70**: 446–51. doi:10.1093/analys/anq033

Elkins, Kathleen (2018) ‘Warren Buffett: "If I had been a female, my life would have been entirely different"'’ CNBC interview, <https://www.cnbc.com/2018/10/16/warren-buffett-says-being-a-white-man-helped-him-succeed.html>

Epstein, Brian (2015) *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford University Press.

Epstude, Kai and Kai J Jonas (2015) ‘Regret and Counterfactual Thinking in the Face of Inevitability: The Case of HIV-Positive Men’, *Social Psychological and Personality Science* **6**: 157–63. doi:10.1177/1948550614546048

Faller, August (2023) ‘Are There Really Social Causes?’, *Philosophy of the Social Sciences* **53**: 83–102. doi:10.1177/00483931221150485

French, Rohan, Patrick Girard, and Ellie Ripley (2022) ‘Classical Counterpossibles’, *Review of Symbolic Logic* **15**: 259–75. doi:10.1017/S1755020319000637

Garry, Ann (2011) ‘Intersectionality, Metaphors, and the Multiplicity of Gender’, *Hypatia* **26**: 826–50. doi:10.1111/j.1527-2001.2011.01194.x

Greenman, Emily, and Yu Xie (2008) ‘Double Jeopardy? The Interaction of Gender and Race on Earnings in the U.S’, *Social forces* **86**:3 1217-1244. doi:10.1353/sof.0.0008

Griffith, Aaron (2018a) ‘Social Construction: Big-G Grounding, Small-g Realization’, *Philosophical Studies* **175**: 241–60. doi:10.1007/s11098-017-0865-x

Griffith, Aaron (2018b) ‘Social Construction and Grounding’, *Philosophy and Phenomenological Research* **97**: 393–409. doi:10.1111/phpr.12376

Haslanger, Sally (2013) ‘Race, Intersectionality, and Method: A Reply to Critics’, *Philosophical Studies* **171**: 109–19. doi:10.1007/s11098-013-0244-1

Hu, Lily (forthcoming) ‘Normative Facts and Causal Structure’, *Journal of Philosophy*.

Jenny, Matthias (2018) ‘Counterpossibles in Science: The Case of Relative Computability’, *Noûs* **52**: 530–60. doi:10.1111/nous.12177

Jenkins, Katharine (2020) ‘Ontic Injustice’, *Journal of the American Philosophical Association* **6**: 188–205. doi:10.1017/apa.2020.6.

Jorba, Marta and Maria Rodó-Zárate (2019) ‘Beyond Mutual Constitution: The Properties Framework for Intersectionality Studies’, *Signs: Journal of Women in Culture and Society* **45**: 175–200. doi:10.1086/703499

Kocurek, Alex (2018) ‘Counteridenticals’, *Philosophical Review* **127**: 323–69. doi:10.1215/00318108-6718783

- Kocurek, Alex (2021) ‘Counterpossibles’, *Philosophy Compass* **16**. doi:10.1111/phc3.12787
- Koslicki, Kathrin (2016) ‘Where Grounding and Causation Part Ways: Comments on Schaffer’, *Philosophical Studies* **173**: 101–12. doi:10.1007/s11098-015-0559-9.
- Kripke, Saul (1980) *Naming and Necessity*. Harvard University Press.
- Le Guin, Ursula K (1969) *The Left Hand of Darkness*. Ace Science Fiction Books.
- Levy, I and S Ben-David (2015) ‘Mechanism of Bystander-Blaming: Defensive Attribution, Counterfactual Thinking, and Gender’, *International Journal of Offender Therapy and Comparative Criminology* **59**: 96–113. doi:10.1177/0306624X13503297
- Lewis, David (1973a) ‘Causation’, *Journal of Philosophy* **70**: 556–67. doi:10.2307/2025310.
- Lewis, David (1973b) *Counterfactuals*. Blackwell.
- Lewis, David (1979) ‘Counterfactual Dependence and Time’s Arrow’, *Noûs* **13**: 455–76. doi:10.2307/2215339
- Lewis, David (2000) ‘Causation as Influence’, *Journal of Philosophy* **97**: 182–97. doi:10.2307/2678389.
- List, Christian and Peter Menzies (2009) ‘Nonreductive Physicalism and the Limits of the Exclusion Principle’, *Journal of Philosophy* **106**: 475–502. doi:10.5840/jphil2009106936
- Malkin, Marc (2019) ‘Elton John Has a Message for Struggling LGBTQ Youth: “Be Proud of Who You Are”’ *Variety* <https://variety.com/2019/music/features/elton-john-on-being-gay-1203248152/>
- Mallon, Ron (2016) *The Construction of Human Kinds*. Oxford University Press.
- Mason, Rebecca and Kate Ritchie (2020) ‘Social Ontology’, in R Bliss and J Miller, eds., *Routledge Handbook of Metametaphysics* Routledge. 312-24. doi:10.4324/9781315112596-24/
- McDonnell, Neil (2018) ‘Transitivity and Proportionality in Causation’, *Synthese* **195**: 1211–29. doi:10.1007/s11229-016-1263-1
- Nolan, Daniel (1997) ‘Impossible Worlds: A Modest Approach’, *Notre Dame Journal of Formal Logic* **38**: 535–72. doi:10.1305/ndjfl/1039540769
- Nolan, Daniel (2013) ‘Why Historians (and Everyone Else) Should Care about Counterfactuals’, *Philosophical Studies* **163**: 317–35. doi:10.1007/s11098-011-9817-z
- Pachter, Lee and Cynthia Coll (2009) ‘Racism and child health: a review of the literature and future directions’, **30** 3:255-63. doi: 10.1097/DBP.0b013e3181a7ed5a.

- Pagano, Emilie (2026) 'Being Social, Being Socially Constructed, and Being Fundamental Relative to Social Reality', *Philosophical Studies* **183**. doi:10.1007/s11098-025-02446-1
- Parfit, Derek (1984) *Reasons and Persons*. Oxford University Press.
- Payton, Dee (2022) 'Searching for Social Properties', *Philosophy and Phenomenological Research* **106**: 741–54. doi:10.1111/phpr.12901
- Pesaran, M Hashem and Ron Smith (2016) 'Counterfactual Analysis in Macroeconometrics: An Empirical Investigation into the Effects of Quantitative Easing', *Research in Economics* **70**: 262–80. doi:10.1016/j.rie.2016.01.004
- Raven, Michael J (2022) 'A Puzzle for Social Essences', *Journal of the American Philosophical Association* **8**: 128–48. doi:10.1017/apa.2020.48
- Rea, Michael C (2022) 'Gender as a Self-Conferred Identity', *Feminist Philosophy Quarterly* **8**. doi:10.5206/fpq/2022.2.13959
- Roth, Veronica (2011) *Divergent*. HarperCollins.
- Rubenstein, Ezra (2024a) 'Proportionality in Causation, Part I: Theories', *Philosophy Compass* **19**: e12957. doi:10.1111/phc3.12957
- Rubenstein, Ezra (2024b) 'Proportionality in Causation, Part II: Applications and Challenges', *Philosophy Compass* **19**: e12960. doi:10.1111/phc3.12960
- Sartorio, Carolina (2010) 'The Prince of Wales Problem for Counterfactual Theories of Causation', in Helen Beebe, Christopher Hitchcock, and Huw Price, eds., *New Waves in Metaphysics*, 259–276. Palgrave Macmillan. doi:10.1057/9780230297425_14
- Savage, Dan (2013) 'What God Wants', *New York Times*, 11 April. <https://www.nytimes.com/2013/04/14/books/review/does-jesus-really-love-me-by-jeff-chu.html>
- Schaffer, Jonathan (2005) 'Contrastive causation', *Philosophical Review* **114** (3):327–358. doi:10.1215/00318108-114-3-327
- Schaffer, Jonathan (2016) 'Grounding in the Image of Causation', *Philosophical Studies* **173**: 49–100. doi:10.1007/s11098-014-0438-1
- Schaffer, Jonathan (2017) 'Social Construction as Grounding; or: Fundamentality for Feminists, a Reply to Barnes and Mikkola', *Philosophical Studies* **174**: 2449–65. doi:10.1007/s11098-016-0738-8
- Shapiro, Eve (2007) 'Drag Kinging and the Transformation of Gender Identities', *Gender and Society* **21**: 250–71. doi:10.1177/0891243206294509

- Shayne, Julie (2020) 'If Trump Were a Woman', *Ms. Magazine*.
<https://msmagazine.com/2020/12/07/woman-president-if-donald-trump-were-a-woman/>
- Silver, Kenneth (2024) 'Backwards Causation in Social Institutions', *Erkenntnis* **89**: 1973–91.
 doi:10.1007/s10670-022-00613-y
- Sinclair, Jim (2013) 'Why I Dislike "Person First" Language', *Autonomy: The Critical Journal of Interdisciplinary Autism Studies* **1**: 2–3.
- Sparrow, Robert (2005) 'Defending Deaf Culture: The Case of Cochlear Implants', *Journal of Political Philosophy* **13**: 135–52. doi:10.1111/j.1467-9760.2005.00217.x.
- Steinem, Gloria (1983) 'If Men Could Menstruate', in *Outrageous Acts and Everyday Rebellions*: 280–281. Holt, Rinehart and Winston.
- Stoljar, Natalie (2018) 'Gender and the Unthinkable', in P Garavaso, ed., *The Bloomsbury Companion to Analytic Feminism*: 123–43. Bloomsbury.
- Tan, Peter (2019) 'Counterpossible Non-vacuity in Scientific Practice', *Journal of Philosophy* **116**: 32–60. doi:10.5840/jphil201911612
- Taylor, Keeanga-Yamahtta (1977/2017) *How We Get Free: Black Feminism and the Combahee River Collective*. Haymarket Books.
- Timpe, Kevin (2020) 'Defiant Afterlife: Disability and Uniting Ourselves to God', in Michelle Panchuk and Michael Rea, eds., *Voices from the Edge: Centering Marginalized Perspectives in Analytic Theology*, 206–232, Oxford University Press.
 doi:10.1093/oso/9780198848844.003.0010
- Touborg, Caroline Torpe (2022) 'Relativizing Proportionality to a Domain of Events', *Synthese* **200**, 91. doi:10.1007/s11229-022-03596-z
- Trogon, Kelly (2013) 'Grounding: Necessary or Contingent?', *Pacific Philosophical Quarterly* **94**: 465–85. doi:10.1111/papq.12009
- Walker, Rebekah, Joni Williams, and Leonard Egede (2016) 'Influence of Race, Ethnicity and Social Determinants of Health on Diabetes Outcomes', *American Journal of the Medical Sciences* **351**: 366–73. doi: 10.1016/j.amjms.2016.01.008
- Weinberger, Naftali (2022) 'Signal Manipulation and the Causal Analysis of Racial Discrimination', *Ergo* **9**: 46. doi:10.3998/ergo.2915
- Wilhelm, Isaac (manuscript) 'Gender Is Essential to Some, but Not All, Individuals'.
- Williamson, Tim (2007) *The Philosophy of Philosophy*. Blackwell.

Wilson, Alastair (2018) ‘Grounding Entails Counterpossible Non-Triviality’, *Philosophy and Phenomenological Research* **96**: 716–28. doi:10.1111/phpr.12305.

Wilson, Alastair (2021) ‘Counterpossible Reasoning in Physics’, *Philosophy of Science* **88**: 1113–24. doi:10.1086/714706.

Witt, Charlotte (2011) *The Metaphysics of Gender*. Oxford University Press.

Yablo, Stephen (1992) ‘Mental Causation’, *Philosophical Review* **101**: 245–80. doi:10.2307/2185535