Philosophy of Science (2024), **91**, 1098–1108
doi:10.1017/psa.2023.166

PSA PHILOSOPHY OF SCIENCE ASSOCIATION

PHILOSOPHY OF SCIENCE

SYMPOSIA PAPER

Causal Agnosticism About Race: Variable Selection Problems in Causal Inference

Alexander Williams Tolbert

Department of Quantitative Theory and Methods, Emory University, Atlanta, GA, USA
Email: alexander.tolbert@emory.edu, https://www.alexanderwilliamstolbert.com/

(Received 20 April 2023; revised 06 October 2023; accepted 12 November 2023; first published online 16 February 2024)

Abstract

This paper proposes a novel view in the the philosophy of race & causation literature known as "causal agnosticism" about race. Causal agnosticism about race implies that it is reasonable to refrain from making judgments about whether race is a cause. The paper's thesis asserts that certain conditions must be met to infer that something is a cause, according to the fundamental assumptions of causal inference. However, in the case of race, these conditions are often violated. By advocating for causal agnosticism, the paper suggests a more modest approach to understanding the role of race in causal relationships.

1. Introduction

The causal status of race has been a subject of debate in social science, computer science, statistics, and causal inference literature (Holland, 1986; Glymour, 1986; Greiner and Rubin, 2011; VanderWeele and Robinson, 2014; Glymour and Glymour, 2014; Sen and Wasow, 2016; Krieger and Davey Smith, 2016; Pearl, 2018; Kohler-Hausmann, 2018; Hu and Kohler-Hausmann, 2020) Much of the debate in the literature over the causal status of race stems from the supposed non-manipulability of race raised by Holland (1986). According to Holland (1986), all causes are manipulable. Because we cannot manipulate race (or at least it is unclear what would go into manipulating it), race must not be a cause. Holland does not specify what it means to be manipulable in the relevant sense here. Nevertheless, this argument has led to much of what I will coin as *the philosophy of race & causation* literature the philosophy of race & causation. The views in *the philosophy of race & causation* the philosophy of race & causation literature can roughly be divided into four main camps:

- "Race is not a cause on non-manipulable grounds" (Holland, 1986).
- "Race is a cause on manipulable grounds" (Pearl, 2018).
- "Race is a cause in some non-manipulable way" (Glymour and Glymour, 2014; Kohler-Hausmann, 2018; Hu and Kohler-Hausmann, 2020).

- "Race is not a cause on non-manipulable grounds, but aspects of race or proxies of race are causal on manipulable grounds" (Greiner and Rubin, 2011; Sen and Wasow, 2016; Weinberger, 2023).

This paper will put examining these views aside and defend the notion of *causal agnosticism* about race, asserting that it is reasonable to withhold judgment on whether race is a cause. It is worth noting that the philosophy of race & causation literature has not made explicit the epistemic distinction as to whether we have good scientific methodological reasons to believe race is a cause as opposed to the metaphysical claim that race is or is not a cause on either interventionist or some other grounds. In the philosophy of race & causation literature, the metaphysical and the epistemic theses are often blurred together without any clear distinction. Presumably, a metaphysical or epistemic justification concerning the philosophy of race & causation may rest on different standards, assumptions, and premises that may not be the same. This paper contributes to this ongoing discussion by making this critical distinction explicit in its central thesis and defending that it is reasonable to be agnostic about the causal effects of race. This thesis is epistemic, not metaphysical. The following sections of this paper will develop and defend the argument for causal agnosticism about race. In section 2, I present the argument schema for causal agnosticism about race. Section 3 defends the first premise, the *causal premise*. Section 4 will argue for the second premise, the *race is confounded* premise, establishing that race has many unmeasured confounders. In section 5, I address potential objections and respond to further strengthen the case for causal agnosticism. Finally, I offer concluding remarks in section 6.

2. Argument for causal agnosticism about race

My argument for *causal race agnosticism* rests on two key premises[1]:

Premise 1 (causal premise):
For all A and Y, if the hypothesis that A causes Y has many unmeasured confounders, then it is reasonable to withhold judgment on whether A causes Y.

Premise 2 (race is confounded):
For all R and Y, if R represents race, then the hypothesis that R causes Y has many unmeasured confounders.

Thesis (causal agnosticism about race):
For all R and Y, if R represents race, it is reasonable to withhold judgment on whether R causes Y.

Regarding confounding, I will narrow my focus to two types of confounding that hinder the estimation of causal effects: *positivity* violations and *ambiguity*, which I describe in more detail in the following sections.

3. Defending the causal premise

The *causal premise* is required for establishing unbiased estimates of the causal effects in the causal inference literature (Pearl, 2009; Hernán and Robins, 2020). The

[1] Thanks to Kareem Khalifa's collaboration in this argument schema.

https://doi.org/10.1017/psa.2023.166 Published online by Cambridge University Press

population average causal effect (ACE) of an intervention on an outcome can be estimated by comparing the hypothetical intervention of everyone receiving the treatment with everyone not receiving it for a well-defined target population. We rely on certain assumptions to estimate the average treatment effect from observational data. When we use the observed outcome distribution of individuals who received an intervention to approximate their hypothetical distribution without the intervention, we actively assume that their observed distribution under the intervention reasonably approximates what we would have observed if they had not undergone the intervention. In simpler terms, we use the outcome distribution of individuals who received the intervention to estimate the distribution of outcomes had the intervention not been given. We assume that individuals who underwent the intervention are significantly comparable to those who did not. We attribute any differences between these groups to other factors, such as demographic characteristics or health status. To achieve this, we assume what is commonly called "exchangeability" between those who receive the intervention and those who do not with respect to the relevant pre-intervention features (Dawid, 2015, p. 282–283).

However, as Dawid (2015) points out, this principle of exchangeability might not apply to the subset of individuals who underwent treatment, as these characteristics could have influenced the decision to administer treatment. For example, a particular medication may have been prescribed only to individuals of a certain group at higher risk for a specific condition. This association may induce a spurious relationship between the treatment and outcome variables, so the actual causal effect is obfuscated. This scenario is commonly known as *confounding*, which hinders interpreting causal relationships in observational data (Dawid, 2015, p. 282-283). More generally, suppose we have some random variables X, A, and Y. To say that A causes Y, we need to rule out the possibility that some other factor, such as X, is responsible for our observed relationship. One way to do this is to measure all the relevant variables and control for them statistically. However, suppose many unmeasured variables could be confounding the relationship. In that case, we cannot be sure that we have ruled out the possibility of alternative explanations by what is referred to as "unmeasured confounding" in the causal inference literature. An unmeasured confounder is a variable connected to treatment and outcome variables that could account for or explain the observed relationship between the two variables (Ananth and Schisterman, 2018, p. 1). Failure to account for all relevant confounding variables can lead to biased treatment effect estimates and invalidate causal inferences. In this case, the causal premise says that it is reasonable to withhold judgment about whether A causes Y.

4. Defending the "race is confounded" premise

4.1. Race and positivity

Recall that failures of positivity are one of two reasons supporting agnosticism. Zivich et al. (2022) identify two types of positivity, *deterministic* and *stochastic* positivity. Deterministic positivity ensures that for every value of the covariates X, there is a non-zero probability of receiving each level of the treatment A. Zivich et al. (2022) mathematically express deterministic positivity as:

https://doi.org/10.1017/psa.2023.166 Published online by Cambridge University Press

$$\Pr(A = a|X = x) \geq \varepsilon > 0 \quad \text{for } a \in \{0, 1\} \text{ and } x \text{ where } p(x) > 0.$$

Here, ε is a small positive number that bounds the probabilities away from zero. The expression $\Pr(A = a|X = x) \geq \varepsilon > 0$ means that for each value of the covariates X, and for each level of the treatment A, the probability of receiving that level of treatment is at least some small positive number. This ensures that every subgroup defined by X has some chance of receiving each level of the treatment A. Violating deterministic positivity can lead to bias by providing inaccurate estimates of the treatment effect for subgroups not exposed to it. In a dataset, specific subgroups may have zero probability of exposure to a particular intervention. Deterministic positivity is violated if a subset of the population has zero probability of receiving treatment, typically due to definitional or inherent structural characteristics. Pharmacoepidemiology often deals with the issue of positivity as drugs are prescribed based on specific conditions such as age, diagnosis, and weight. This results in frequent deterministic positivity violations in such settings (Platt et al., 2012). Another typical example of deterministic non-positivity is occupational exposure. Once workers fall ill due to chemical exposure at the workplace, they are usually sent home or to the hospital, leading to a zero chance of exposure after that point (Naimi et al., 2011).[2]

On the other hand, stochastic positivity concerns the probability of treatment in a particular sample drawn from the population. It ensures that every possible sample of individuals has some chance of receiving both levels of the treatment. Even if every individual in the entire population has a non-zero probability of receiving the treatment, it is still possible that by random chance, a sample might be drawn in which a subgroup has zero probability of receiving the treatment. Zivich et al. (2022) mathematically express stochastic positivity as:

$$\Pr_n(A = a|X = x) > 0 \quad \text{for } a \in \{0, 1\} \text{ and } x \in \{x_1, x_2, \ldots, x_n\}.$$

Here, $\Pr_n$ denotes the conditional probability in the sample of size n. The expression $\Pr_n(A = a|X = x) > 0$ implies that in every possible sample of individuals from the population, there is a positive probability of receiving each level of the treatment A for every value of the covariates X. Unlike deterministic positivity, this is about the sampling process, not a structural or definitional impossibility based on covariate values. This ensures that the treatment effect can be estimated in all possible instances, regardless of their characteristics. Stochastic non-positivity is a finite sampling issue due to the inherent variability in data collection. Deterministic positivity and stochastic positivity are crucial concepts that ensure the validity of causal inference.

With positivity explained, the rest of this section will show that race can violate positivity, creating confounding. In the social sciences, several studies have remarked on the confounding of race with socioeconomic status (LaVeist, 2005; LaVeist et al., 2007). For example, consider race and socioeconomic status (SES). The work of Messer et al. (2010), discussed by VanderWeele and Robinson (2014, p. 477), highlights the difficulty in differentiating the effects of SES and race. They point out that in situations characterized by significant income disparities, where individuals from a

[2] Thanks to Jay Kaufman for these references and examples.

https://doi.org/10.1017/psa.2023.166 Published online by Cambridge University Press

specific racial group exclusively occupy a particular SES, it becomes difficult to disentangle the influences of SES and race. This difficulty persists even when data on these variables are available. LaVeist (2005, p. iii27) also points out how the absence of overlap between race and SES complicates isolating each variable's causal influence in the context of health disparities research. Racial minorities tend to occupy lower socioeconomic tiers than Whites, complicating the analysis of whether health disparities arise from race and social class together or race or social class separately. Hence, disentangling the effects of race and socioeconomic status in health inequities proves challenging. In the context of SES and race, positivity violations can result from the structural patterns created by segregation and racism. Segregation and racism structure the data in a way that highly correlates with race and poverty in specific areas of the United States (Messer et al., 2010). However, it is essential to note that these patterns are not deterministic, meaning there are exceptions to the general trend. For example, there are affluent Black neighborhoods and poor White neighborhoods, although these instances may be relatively sparse in specific geographic settings. Nevertheless, when race is an exposure, non-positivity will exist whenever there is extreme social stratification, thus becoming a problem epistemically for causal inference.

The absence of individuals from a marginalized group in privileged societal positions leads to substantial positivity violations, some of which may even be deterministic. This situation arises when discriminatory practices and social structures systematically prevent marginalized group members from attaining positions of power, privilege, or access to resources. For example, historical instances like the caste system in India, the antebellum South in the United States, or the occupation of Ireland by the British exemplify extreme discrimination that resulted in positivity violations. In these cases, the discriminatory systems were deeply entrenched, creating structural barriers that limited the opportunities for individuals from marginalized groups to advance socioeconomically or gain access to higher social positions.

For example, in the U.S., slavery and racial segregation were widespread. Enslaved and marginalized African Americans were denied education, economic opportunities, and political power. The systemic discrimination and the institution of slavery ensured that individuals from the enslaved population could not occupy privileged positions in society. Furthermore, economists Hamilton and Darity, Jr. (2010) argue that the broken promise of 40 acres and a mule to ex-slaves, coupled with property deprivation of Black Americans between 1880 and 1910, has led to a racial wealth gap in the U.S. This gap is perpetuated by structural barriers fueled by past and present discrimination. Inheritances, bequests, and intra-family transfers contribute more to this gap than education and income, with White families receiving larger estates on average than African-American families. As a result, there was and still is a lack of overlap between African Americans and White individuals because of a history of discrimination. More generally, extreme discrimination will always limit the overlap between groups, contributing to the absence of certain combinations or subgroups in the data, leading to positivity violations.

Further, conceptual dilemmas arise for some views of race. For example, Sally Haslanger's social construction view of race is defined as racialized practices constructing social realities around physical features. In her account, "hierarchical

https://doi.org/10.1017/psa.2023.166 Published online by Cambridge University Press

positioning of an ethnic group within a broader society (or broader political formation) is a process of racializing the group" (Glasgow et al., 2019, p. 27). I refer to this as Haslanger's racialization thesis. Haslanger claims that racialization is the process by which social races are formed (Glasgow et al., 2019). The racialization thesis creates a dilemma in knowing the causal effects of *race* and the metaphysical thesis of social races. The first horn of the dilemma is that accepting racialization leads to positivity violations due to the lack of overlap between racialized groups. Positivity violations hinder causal inferences about the effects of race. The inability to make causal inferences undermines the knowledge claim about racialization's role in forming and maintaining social races. This leads to a tension where accepting racialization undermines the very claim about racialization. On the second horn of the dilemma, suppose we reject racialization to ease causal inference. However, this contradicts the racialization thesis, and as a result, this rejection undermines the metaphysical claim about the existence of social races. In other words, easing causal inference comes at the cost of denying the metaphysical reality of social races formed through racialization. To know the causal effects of race, one has to deny the metaphysical reality of social races.

Tension arises when a group is increasingly subjected to racialization, leading to heightened discrimination against them. As a result, there will not be any members of that group in privileged positions in society, creating a lack of overlap between the groups and a violation of positivity. This is what makes the dilemma apparent. If Hasslangerian social constructivists hold that racialization is accurate, the more they believe racialization holds, the less they can be sure about race's causal effects because of positivity violations. However, if they deny the force of racialization in our current society to know race's effect, they deny the existence and maintenance of race in Hasslanger's terms. They must trade knowledge claims about race's causal powers with metaphysical claims about race's existence as a social construction according to how they conceive it. The dilemma leads to undermining our knowledge about the reality of social races. Either accepting or rejecting racialization challenges the ability to affirm the knowledge about the reality of social races.

One implication from my argument is that to solve our epistemic problem we may need to solve our moral one first. Hamilton and Darity, Jr. (2010), mentioned earlier, focused on normative grounds for reparations for reasons rooted in the racial wealth gap. However, my arguments suggest that we also have epistemic reasons that can serve as grounds for repair. Fixing the social-political problem of social inequality among groups contributing to the lack of positivity would make causal inference easier. This epistemic conundrum may open another route for generating normative obligation for reparations and other forms of social egalitarianism.

4.2. Race and ambiguous variables

In the causal inference literature discourse, an ambiguous treatment variable is identified as one whose manipulation encompasses multiple underlying causal variables, the proportions of which are either unclear or inconsistent, leading to uncertain or inconsistent outcomes. The ambiguity lies in the underdetermination of underlying causal variables' values when the ambiguous variable's value is specified. This underdetermination obfuscates the causal inferences that can be drawn from

https://doi.org/10.1017/psa.2023.166 Published online by Cambridge University Press

manipulating the ambiguous variable on the outcome of interest (Spirtes and Scheines, 2004, p. 834).

Consider a scenario wherein researchers are probing the effects of a "15 minutes of daily exercise" regimen on stress reduction. The recommendation comes from the hypothesis that a daily exercise routine mitigates stress levels. However, this variable is constitutive of aerobic and anaerobic exercises. The former, consisting of activities like jogging or swimming, is known to ease stress. In contrast, the latter, including activities like weightlifting, may either have a negligible effect or exacerbate stress in some individuals. Furthermore, there are varying intensities in which these exercises can be conducted. The ambiguous variable here, "15 minutes of daily exercise," does not specify the type nor nature of exercise, thus leading to a range of outcomes in stress reduction. The ambiguity is rooted in the underdetermination of the type and intensity of exercise when the regimen of "15 minutes of daily exercise" is prescribed.[3]

Woodward (2016) notes that many variables lack stability due to the possibility of ambiguous manipulations. Scholars such as Spirtes and Scheines (2004) and Woodward (2016) have both discussed total cholesterol (TC) as a paradigmatic example in this context. Woodward (2016) explains that research on the effect of cholesterol on heart disease D measured TC treated as the sum of low-density lipoprotein (LDL) and high-density lipoprotein (HDL), and revealed that LDL and HDL have very different effects on D. Higher levels of LDL increase the probability of D. In contrast, higher levels of HDL decrease the probability of D. The effect of an intervention that sets TC to a specific value v on D is ambiguous, as it depends on the precise mix of LDL and HDL involved in this particular realization of $\mathrm{TC} = v$ (Woodward, 2016, p. 1069). To clarify the point, consider an observational study where researchers determine a correlation between high levels of TC and D. An intervention for a diet low in cholesterol may vary in the proportions of LDL and HDL cholesterol they affect, thereby leading to a divergence in outcomes in experiments with low-cholesterol regimens. Here, the ambiguous variable is TC, the manipulation of which affects the levels of LDL and HDL cholesterol in varying proportions, thereby obscuring effects on heart disease. The ambiguity stems from the underdetermination of LDL and HDL cholesterol values when the value of total cholesterol is specified, leading to a lack of clarity in causal inference regarding heart disease prevention.

Similar problems arise with the cluster-kind views of race. By cluster-kind views of race, I mean a theory that posits race as a higher-order macrovariable constitutive of lower-order microvariables. Social science has cluster-kind accounts of race that posit race is constitutive of many variables. For instance, Sen and Wasow (2016, p. 506) suggest a variety of potential variables that could be considered, including ancestry region, wealth, dialect, genetic factors, neighborhood characteristics, diet, social standing, norms, power dynamics, class, skin color, religion, and region of origin. Consider clinical trials aiming to establish drug efficacy; the variable "race" often emerges as a significant factor. However, suppose race is constitutive of self-identification and genetic ancestry, each bearing distinct genetic variabilities that could significantly impact the drug's effectiveness. Researchers might recruit a more uniform pool based on actual genetic ancestry to minimize genetic variability within a

[3] This is related to "no multiple versions of treatment," an assumption needed for causal inference included in the stable-unit-treatment-value assumption; see Hernán and Robins (2020, p. 6).

https://doi.org/10.1017/psa.2023.166 Published online by Cambridge University Press

racial group and thereby better isolate the drug's effects. While reducing genetic variability, this method might inadvertently select prevalent genetic traits within this homogeneous group that result in poor drug metabolism or adverse reactions, making the drug appear ineffective. Conversely, increasing the representation of a given race by self-identification might lead to a more genetically diverse group. This genetic diversity could potentially cast the drug in a more favorable light if the genetic diversity includes traits that respond well to the drug. However, this approach does not account for the inherent genetic variability, which could lead to an overestimation of the drug's effectiveness.

Further, substantial heterogeneity within racial categories can impede reliable inference. A salient example is using the "Black" category in income disparity analyses. The "Black" category aggregates individuals from a wide range of ethnic and ancestral backgrounds, like those tracing their roots to various regions of Africa, Afro-Caribbean individuals, or those residing in the United States for multiple generations. Using "Black" as an aggregate variable combines diverse experiences, backgrounds, and social factors within this category. Consider the contrasting experiences of a recent Nigerian immigrant and a U.S. resident descended from enslaved Africans within the "Black" category. Recent Nigerian immigrants, potentially benefiting from selective immigration policies, may attain higher levels of education, allowing access to lucrative professional opportunities and potentially resulting in higher income (Sakamoto et al., 2021).

On the other hand, descendants of enslaved Africans in the U.S. grapple with a different socioeconomic landscape. As mentioned earlier, the legacy of slavery, unfulfilled restitution promises, and systemic property deprivation have shaped their economic reality, creating a racial wealth gap in the U.S. and affecting African Americans' access to education. These factors collectively relegate many within this subgroup to lower-income strata. This heterogeneity is often obscured within the "Black" category, making the causal relationships with income ambiguous. In this context, disaggregated microvariables, such as specific ancestry, immigration history, and even wealth, may offer more stable and, thus, better income predictors than the macrovariable. These variables might present more stable effects on income, enabling a clearer understanding of the causal relationships involved. Note that microvariables are not "race" itself that we would know the effects about but rather constitutive components of race. This means we need to withhold judgment about the macrovariable, only knowing the effects of its microvariables. More generally, because race is an ambiguous macrovariable, it introduces multiple unmeasured confounders, which suggests withholding judgment about its causal effects.

This highlights the importance of stable causal relationships and understanding the appropriate level of aggregation for variables. Woodward (2016) argues that stability is a condition that should constrain variable choice. For Woodward (2016), stability refers to the degree to which a causal relationship between two variables, X and Y, is generalizable from one set of circumstances to another. I take stability to refer to the uniformity of a conditional distribution across different data realizations. This means that the conditional distribution of Y given X and the non-stochastic regime indicator F_X, which indexes the different conditions under which a system is observed, remains the same regardless of the specific conditions or state under which

https://doi.org/10.1017/psa.2023.166 Published online by Cambridge University Press

the data is generated.[4] Mathematically, this can be expressed with the conditional independence property, denoted as $Y \perp\!\!\!\perp F_X | X$. This property suggests that the probability distribution of Y, conditioned on both the value of X and the regime F_X that led to that value, remains consistent across all possible regimes.

5. Objections and replies

5.1. Objection

This section addresses an objection to the claim that race violates the positivity assumption. The objection argues that certain variables, such as SES or neighborhood, should not be included in the conditioning set because they are post-treatment variables that are descendants of racialized status. This objection asserts that including post-exposure variables in the conditioning set would block essential pathways through which race affects outcomes and would not violate positivity. A post-treatment variable is a variable that is measured or observed after an individual has been exposed to a particular intervention of interest and is often used in causal inference to evaluate the effects of the treatment on a particular outcome of interest. Recall that the positivity assumption is formally stated as $P(A = a|X) > 0$, where X is the set of variables sufficient to adjust for identifying the ACE. These Xs are confounders, not mediators or anything post-exposure (descendants of A would not be in X).

The variables that should be included in X depend on what might confound the A, Y relationship and factors influencing racialized group membership and the outcome. For example, age and gender might be included because age is correlated with race by various mechanisms, and gender plausibly affects racialization according to intersectionality theory. However, $P(A = a|\text{age}, \text{gender}) > 0$ should not pose a problem in any reasonable study design. However, variables like SES or neighborhood are clearly "post-exposure" because they are causally descendants of racialized status, not the reverse. The objection claims that a good study design would never include them in X.[5]

5.2. Reply

Although SES and wealth are typically considered post-exposure variables downstream of racialized status, the racialization thesis posits that these variables also play a role in the formation and maintenance of race. Therefore, while it is true that controlling for these variables may block one pathway through which race affects outcomes, according to the racialization thesis, these variables contribute to the social context in which race operates and is formed. According to racialization, we should hold that something like an "oppression" macrovariable has a causal arrow flowing into race. If we were to ask ourselves what this variable was constitutive of, it would likely reasonably be things like SES, income, neighborhood, and wealth. The objection fails to recognize that variables, such as wealth and SES, are not just downstream effects of race but also contribute to the formation and maintenance of race itself. In other words, the causal relationship is cyclic, and these variables have

[4] For details on regime indicators, see Dawid (2015).

[5] Thanks to Daniel Malinsky for raising this insightful objection.

https://doi.org/10.1017/psa.2023.166 Published online by Cambridge University Press

causal arrows going into race. Furthermore, to deny this would be to deny racialization and, thus, social races formed via racialization.

6. Conclusion

The argument for causal race agnosticism rests on two crucial premises: the "causal premise" and the notion that "race is confounded." According to the causal premise, when many unmeasured factors confound the observed relationship between two variables, it is reasonable to withhold judgment about their causal relationship. Certain conditions must be met to make causal inference possible, but these conditions are often violated in the case of race. For instance, structural racism guarantees that race will highly correlate with other variables like poverty, which violates positivity. Additionally, macrovariable views of race often violate the no-ambiguous-treatment-variable principle. These issues pose significant challenges to inferring causal conclusions about race's effects. Therefore, embracing causal agnosticism about race is a more modest epistemic approach.

References

Ananth, Cande V. and Enrique F. Schisterman. 2018. "Hidden Biases in Observational Epidemiology: The Case of Unmeasured Confounding". *BJOG* 125 (6):644–646. doi: 10.1111/1471-0528.14960

Dawid, A. Philip. 2015. "Statistical Causality from a Decision-Theoretic Perspective". *Annual Review of Statistics and Its Application* 2 (1):273–303. doi: 10.1146/annurev-statistics-010814-020105

Glasgow, Joshua, Sally Haslanger, Chike Jeffers, and Quayshawn Spencer. 2019. *What is Race? Four Philosophical Views*. Oxford: Oxford University Press.

Glymour, Clark. 1986. "Statistics and Causal Inference: Comment: Statistics and Metaphysics". *Journal of the American Statistical Association* 81 (396):964–966. doi: 10.2307/2289067

Glymour, Clark and Madelyn R. Glymour. 2014. "Commentary: race and sex are causes". *Epidemiology* 25 (4):488–490. doi: 10.1097/EDE.0000000000000122

Greiner, D. James and Donald B. Rubin. 2011. "Causal Effects of Perceived Immutable Characteristics". *The Review of Economics and Statistics* 93 (3):775–785. doi: 10.1162/REST_a_00110

Hamilton, Darrick and William Darity, Jr. 2010. "Can 'Baby Bonds' Eliminate the Racial Wealth Gap in Putative Post-Racial America?" *The Review of Black Political Economy* 37 (3–4):207–216. doi: 10.1007/s12114-010-9063-1

Hernán, Miguel and Jamie Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hill/CRC.

Holland, Paul W. 1986. "Statistics and Causal Inference". *Journal of the American Statistical Association* 81 (396):945–960. doi: 10.2307/2289064

Hu, Lily and Issa Kohler-Hausmann. 2020. "What's Sex Got To Do with Machine Learning?" Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. 513.

Kohler-Hausmann, Issa. 2018. "Eddie Murphy and the Dangers of Counterfactual Causal Thinking about Detecting Racial Discrimination". *Northwestern University Law Review* 113 (3):1163.

Krieger, Nancy and George Davey Smith. 2016. "The Tale Wagged by the DAG: Broadening the Scope of Causal Inference and Explanation for Epidemiology". *International Journal of Epidemiology* 45 (6):1787–1808. doi: 10.1093/ije/dyw114

LaVeist, Thomas A. 2005. "Disentangling Race and Socioeconomic Status: A Key to Understanding Health Inequalities". *Journal of Urban Health* 82 (2 Suppl 3):iii26–iii34. doi: 10.1093/jurban/jti061

LaVeist, Thomas A., Roland J. Thorpe, Jr., GiShawn A. Mance, and John Jackson. 2007. "Overcoming Confounding of Race with Socio-Economic Status and Segregation to Explore Race Disparities in Smoking". *Addiction* 102 (Suppl 2):65–70. doi: 10.1111/j.1360-0443.2007.01956.x

Messer, Lynne C., J. Michael Oakes, and Susan Mason. 2010. "Effects of Socioeconomic and Racial Residential Segregation on Preterm Birth: A Cautionary Tale of Structural Confounding". *American Journal of Epidemiology* 171 (6):664–673. doi: 10.1093/aje/kwp435

https://doi.org/10.1017/psa.2023.166 Published online by Cambridge University Press

Naimi, Ashley I., Stephen R. Cole, Daniel J. Westreich, and David B. Richardson. 2011. "A Comparison of Methods to Estimate the Hazard Ratio Under Conditions of Time-varying Confounding and Nonpositivity". *Epidemiology* 22 (5):718–723. doi: 10.1097/EDE.0b013e31822549e8

Pearl, Judea. 2009. *Causality*, 2nd ed. Cambridge: Cambridge University Press.

Pearl, Judea. 2018. "Does Obesity Shorten Life? Or Is It the Soda? On Non-Manipulable Causes". *Journal of Causal Inference* 6 (2):20182001. doi: 10.1515/jci-2018-2001

Platt, Robert W., Joseph A. C. Delaney, and Samy Suissa. 2012. "The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding". *European Journal of Epidemiology* 27 (2):77–83. doi: 10.1007/s10654-011-9637-7

Sakamoto, Arthur, Ernesto F. L. Amaral, Sharron X. Wang, and Courtney Nelson. 2021. "The Socioeconomic Attainments of Second-Generation Nigerian and Other Black Americans: Evidence from the Current Population Survey, 2009 to 2019. *Socius* 7. doi: 10.1177/23780231211001971

Sen, Maya and Omar Wasow. 2016. "Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics". *Annual Review of Political Science* 19 (1):499–522. doi: 10.1146/annurev-polisci-032015-010015

Spirtes, Peter and Richard Scheines. 2004. "Causal Inference of Ambiguous Manipulations". *Philosophy of Science* 71 (5):833–845. doi: 10.1086/425058

VanderWeele, Tyler J. and Whitney R. Robinson. 2014. "On the Causal Interpretation of Race in Regressions Adjusting for Confounding and Mediating Variables". *Epidemiology* 25 (4):473–484. doi: 10.1097/EDE.0000000000000105

Weinberger, Naftali. 2023. "Signal Manipulation and the Causal Analysis of Racial Discrimination". *Ergo* 9:46. doi: 10.3998/ergo.2915

Woodward, James. 2016. "The Problem of Variable Choice". *Synthese* 193:1047–1072. doi: 10.1007/s11229-015-0810-5

Zivich, Paul N., Stephen R. Cole, and Daniel Westreich. 2022. "Positivity: Identifiability and Estimability". Preprint, arXiv:2207.05010.

Cite this article: Tolbert, Alexander Williams. 2024. "Causal Agnosticism About Race: Variable Selection Problems in Causal Inference." *Philosophy of Science* 91 (5):1098–1108. https://doi.org/10.1017/psa.2023.166