

Race Causality Discrimination – Testing for Discrimination

Marcello Di Bello - ASU - Fall 2023 - Week #11

Last week we reviewed how quantitative social scientists measure racial discrimination.¹ They often rely on a counterfactual test: had the person been of a different race, would they have been treated differently? We also looked at some complications of this approach: depending on how the underlying causal model is set up, the counterfactual test may give rise to different verdicts.²

Today we examine more closely whether the counterfactual approach is adequate for studying racial discrimination from a descriptive and normative standpoint.

Kohler-Hausmann

Kohler-Hausmann argues that that, if race is understood according to social constructionism (as many social scientists do), a counterfactual approach to racial discrimination—e.g., using audit or field studies—is inadequate.^{3 4}

The dominant view about race among social scientists is social constructionism.⁵ Race is understood as a social fact that emerges from complex historical and social processes:

the term “race” . . . references a complexly constituted social fact, whereby material and dignitary opportunities are organized such that certain physical and cultural signifiers become the salient markers of consequential cultural categories, and those categories are constituted by a constellation of social relations and meanings with a definite content and organization. (p. 1170)

A thought experiment helps to see why the counterfactual account of racial discrimination is inadequate:

Consider an island society where the categories of social stratification are binary: Royal and non-Royal. The privileged class, Royal, wears purple capes and carries sticks. Their cultural tastes define what is deemed high and valuable in this society, they occupy more prestigious occupations on the island, hoard more resources, etc . . . Royal is a cultural category of thought and action constituted by a complex set of social relations and meanings. (p. 1179-1180)

Suppose an anthropologist visits the island. They do not know anything about the social stratification. They merely observe that those wearing capes and carrying sticks are given precedence on sidewalks. By running an audit study, they can confirm that there is a counterfactual dependence between the variables “attire” and “sidewalk

¹ See National Academies of Sciences, Engineering, and Medicine (2004), *Measuring Racial Discrimination*, The National Academies Press, especially Chapter 5 and the Executive Summary.

² See Barocas, Hardt and Narayanan (2023) *Causality*, especially section “Counterfactuals in the law”, Chapter 5 of *Fairness and Machine Learning*, MIT Press.

³ Inadequate for doing what, exactly? This will be crucial in assessing Kohler-Hausmann’s argument.

⁴ Kohler-Hausmann (2019), Eddie Murphy and the Dangers of Counterfactual Causal Thinking about Detecting Racial Discrimination, *Northwestern University Law Review* 113: 1163-1227

⁵ See previous reading for this class: Haslanger (2019), Tracing the Sociopolitical Reality of Race, in *What is Race? Four Philosophical Views*, Oxford University Press and Jeffers (2019), Cultural Constructionism, in *What is Race? Four Philosophical Views*, Oxford University Press.

behavior”.⁶ Kohler-Hausmann notes that, having established such counterfactual dependence, the anthropologist still has not understood much of what is going on in the island:

The visiting anthropologist could not classify the sidewalk action as non-Royal debasement simply by asking, “If I changed an isolated trait about a person (cape, stick) and nothing else changed about that person, would other pedestrians have remained on the sidewalk?” ... It is only possible for those in the culture to react to the category of Royal because they recognize the indices of Royal to mean something significant beyond the holding of a stick or wearing of a cape. (p. 1180)

The anthropologist failed to understand the underlying “social mechanism” by which people give Royals the right of way on sidewalks. Similarly, audit studies do not give us information about how racial discrimination operates. They merely establish a counterfactual dependence between a behavior and a variable that is associated with race (name, physical appearance, etc.).

As a methodological point, audit studies assume that we can isolate the effects of race by controlling for all other variables. As quasi-randomized control studies, they accomplish that. But there is a problem:

Most audit studies proceed by trying to make all factors about the auditors that might be theoretically relevant to the decision-maker’s decision identical... But ... formally identical credentials are interpreted differently with differently raced candidates precisely because of the social fact of race. (p. 1211)⁷

What we need—Kohler-Hausmann argues—is a *constitutive explanation*⁸, not relations of counterfactual dependence between isolated variables:

To ask if something happened “because of race” similarly calls for a constitutive explanation, one that references the complex system of social meanings and relations that make up the very category. To say, for example, “I was not given a traffic ticket when I was pulled over for speeding because I am white” is to offer a constitutive explanation that references what constitutes WHITENESS—namely, a social type entailing a presumption of noncriminality and deservedness of leniency and respect.

Audit studies are inadequate for another reason: by themselves, they do not provide evidence of discrimination because discrimination is a *thick ethical concept*.⁹

racial discrimination is a thick ethical concept that rests on an account of the system of social meanings that constitute race *and* a normative theory for why (and when) decisions that are based on those social meanings are worthy of moral concern. (p. 1214)

⁶ How would that study be set up?

⁷ Here is an example: “the history of racial segregation and state neglect of black schools has created large differences in the mean quality of schools between white and black neighborhoods in most major cities, so a decision-maker might treat the high school diploma signal differently for a black and white auditor based on this knowledge or presumption.” (p. 1211)

⁸ This is the first conceptual tool that Kohler-Hausmann uses to understand racial discrimination. How is a constitutive explanation defined?

⁹ This is the second conceptual tool that Kohler-Hausmann uses to understand racial discrimination. How is the notion of a thick ethical concept defined? Why does it require an appeal to both descriptive and normative concepts?

Discrimination understood as counterfactual dependence would regard discrimination *because of* freckles and discrimination *because of* race as formally the same. They both would count as irrational.¹⁰ But to view racial discrimination as irrationality fails to understand it.¹¹

The reason why audit studies provide powerful evidence of discrimination is that they presuppose widespread moral intuitions about the wrongfulness of discrimination:

The reason that most of us recognize results from audit studies showing differential treatment by decision-makers of differently raced candidates is because there is widespread agreement that differently raced persons with the specific credentials presented in the study *ought to be* treated similarly. (p. 1216)

Audit studies are still important for studying discrimination. They help to build ‘a case’ and vouch for a ‘particular constitutive claim with moral dimension’ (p. 1224):¹²

We can still seek to detect discrimination using audit studies, regression techniques with observational data, and many of the same methods folks have long used in social science and legal challenges . . . what we are *not* doing with those methods is detecting the treatment effect of race in the counterfactual causal sense. (p. 1223)¹³

Weinberger

Weinberger disagrees.¹⁴ He draws a distinction between understanding what racial discrimination is (and why it is wrong) and detecting discrimination. He thinks that audit studies cannot do the former, but they can do the latter:

Kohler-Hausmann conflates the issues of whether causal counterfactuals provide an analysis of what discrimination is and why it is bad with the issue of whether such counterfactuals are relevant to *detecting* discrimination, which is what she claims to be addressing. (p. 10)¹⁵

Weinberger’s proposal is that the treatment variable in audit studies (name, appearance, etc.) is a *signal* for race, and this is good enough for detecting (albeit not explaining) discrimination:

In testing whether Non-Royals defer to Royals, one does not manipulate whether someone is a Royal, but whether someone has a cape and stick. This test only works, however, if the cape and stick are reliable signals for royalty. (p. 11)

Weinberger’s signal-manipulation models consists of three variables:

1. C is the category (e.g., race)
2. S is the signal (e.g., name)

¹⁰ “the entire logic of the CCM is to define discrimination as a form of irrationality, which requires units (persons, neighborhoods, etc.) to present with identical functionally relevant characteristics in order to elicit differential effects.” (p. 1186)

¹¹ Can you explain this point more fully? Kohler-Hausmann writes: “There is no reason above and beyond opposition to idiosyncratic aversions, irrationality, random meanness, or a general opposition to the structure of disadvantages (as opposed to whom and how they are allocated) to care about a decisionmaker denying an opportunity or imposing a cost on a freckled candidate. But there is a reason above and beyond opposition to idiosyncratic aversions, irrationality, or a general opposition to the structure of disadvantages to care about a decisionmaker denying an opportunity or imposing a cost on a black candidate.” (p. 1214)

¹² What is the ‘constitutive claim with moral dimension’ that Kohler-Hausmann has in mind here?

¹³ Can you spell out the distinction between *detecting discrimination* and *detecting the treatment effect* of race?

¹⁴ Weinberger (2023), *Signal Manipulation and the Causal Analysis of Racial Discrimination*, *Ergo an Open Access Journal of Philosophy* 9(46): 1264-1287.

¹⁵ Do you think this objection is a fair one? More specifically, in terms the Royal/non-Royal example by Kohler-Hausmann, Weinberger writes: “Intervening to give a Non-Royal a cape and stick and seeing if other Non-Royals defer to them would not tell the researcher why Royals are privileged within the society. But it is a perfectly good test of whether Non-Royals are being discriminated against in the particular context. Such a test would not resolve the question of whether such discrimination is bad or what makes it so, and thus would be of limited use to a lawmaker deciding whether to design legal protections against it. But it would nevertheless be a vital test to conduct if one was uncertain whether Royals and Non-Royals were being treated differently on sidewalks.” (p. 10)

3. Y is the effect to be measured (e.g., callback)

The model assumes that S is a reliable signal for the category C of interest. What is manipulated is the signal S , not the category C itself. Manipulating S is a peculiar kind of intervention:

A signal-based manipulation involves manipulating S . But unlike with ideal interventions, one does not do so to render it fully independent of its prior causes. Rather, one manipulates it to make it vary in the way it would were C to have one value as opposed to another. So if C has two values, c and c' , which are reliably tracked by the values s and s' , respectively, then signal-based manipulations for the influence of C on Y would correspond to the result of manipulating S from s to s' . (p. 11)¹⁶

Beside Kohler-Hausmann's argument, Weinberger also criticizes two other accounts in the literature: (a) replace the 'race' variable with the variable 'perception of race' (Greiner and Rubin); and (b) viewing race as an unstructured bundle of sticks (Sen Wasow). Here is a summary of his criticisms:

Greiner and Rubin focus only on the signal. Sen and Wasow cannot account for the systematic relationship. Kohler-Hausmann recognizes the complexity and non-manipulability of status, but denies that the effects of status are tested via the manipulation of signals. (p. 11)

The signal-manipulation account circumvents some of the objections Kohler-Hausmann levels against the counterfactual account.¹⁷ But a final objection is hard to handle. Consider a case in which changing the signal, while keeping all else the same, affects the other variables. For example, suppose male and female candidates go to an interview wearing the same dress. The meaning of 'dress' is different for male and female interviewees because of gender confirming norms.¹⁸ Weinberger notes:

there exists a spectrum of scenarios ranging from the cases where it is most feasible to isolate an effect of race via manipulation of a signal to those in which race operates in such a holistic manner that such manipulations are no longer coherent. (p. 18)

For Weinberger, the signal-manipulation account allows to conceptualize the causal effects of race in a large spectrum of cases, except fringe cases in which race should not be viewed in causal terms.¹⁹

¹⁶ "In selecting the contrastive values to which one intervenes to set the signal, one assigns the values that would obtain given different instantiations of category C . In this sense, the signal behaves as if it were tracking C . Nevertheless, it is not tracking C (within the experiment), since its variation is not due to variation in C . Although one changes the name on the resume from a black-sounding to a white-sounding name, the applicant remains black." (p. 12)

¹⁷ Can you elaborate on this point?

¹⁸ On such examples, recall Hu (2022), *Causation in the Social World*. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences, in particular Chapter 1 and 2.

¹⁹ What if *all* cases in which race is implicated are like the gender non-conforming example by Lily Hu? If social constructionism about race is correct, couldn't it be that 'race operates in such a holistic manner' *all* the time?