Causally Interpreting Intersectionality Theory

Liam Kofi Bright, Daniel Malinsky, and Morgan Thompson*†

Social scientists report difficulties in drawing out testable predictions from the literature on intersectionality theory. We alleviate that difficulty by showing that some characteristic claims of the intersectionality literature can be interpreted causally. The formalism of graphical causal modeling allows claims about the causal effects of occupying intersecting identity categories to be clearly represented and submitted to empirical testing. After outlining this causal interpretation of intersectional theory, we address some concerns that have been expressed in the literature claiming that membership in demographic categories can have causal effects.

1. Introduction. Intersectionality theory focuses on the idea that people occupy multiple demographic categories. This introduces complexities into social analysis. Work that does not pay explicit attention to these complexities often thereby distorts and misrepresents people's experiences. As a popular slogan has it, social theory concerning gender and race has tended to proceed as if "all the women are white, all the blacks are men." Minimally construed, intersectionality theory is the attempt to correct these analytical failings by directing theorists' attention to the ways in which intersecting demographic categories produce distinctive effects.

Received January 2015; revised May 2015.

*To contact the authors, please write to: Liam Kofi Bright, Philosophy Department, Carnegie Mellon University, Pittsburgh, PA 15213; e-mail: lbright@andrew.cmu.edu. Daniel Malinsky, Philosophy Department, Carnegie Mellon University, Pittsburgh, PA 15213; e-mail: malinsky@cmu.edu. Morgan Thompson, History and Philosophy of Science Department, Pittsburgh University, Pittsburgh, PA 15260; e-mail: mot14@pitt.edu.

†The authors would like to thank Clark Glymour, Mazviita Chirimuuta, Tommy J. Curry, David Danks, Danielle Wenner, Jim Woodward, and participants at the University of California, Irvine Hypatia Conference in 2014, as well as anonymous reviewers for helpful comments.

Philosophy of Science, 83 (January 2016) pp. 60–81. 0031-8248/2016/8301-0003$10.00

Recently, a number of social scientists have called for a better understanding of the predictions of intersectional theory and how to test intersectional hypotheses (e.g., McCall 2005; Bowleg 2008; Cole 2009). Despite the growing popularity of intersectionality theory, discussion of methods of testing intersectional hypotheses has often remained vague. Few works in intersectional theory provide concrete examples or suggestions for empirical tests based on detailed theoretical analysis (Dubrow 2008). Even where concrete examples of the application of intersectional theory have been given, it has still remained unclear how exactly social scientists might test the predictions that follow from these cases (Kantola and Nousiainen 2009). As we shall now show, this has led to much conflicting advice as to how best to test the claims of intersectionality analysis in the social scientific literature.

Difficulties have arisen with attempts to study intersectional hypotheses by quantitative means. The "best-practices guide" for psychologists interested in applying intersectional theory suggests that factorial designs may not be ideal for testing intersectional hypotheses (Warner 2008). Cole (2009) suggests that statistical models may be indispensable for intersectional analyses of employment, income, health, and social life. Yet she cautions that quantitative analyses might miss qualitative differences between groups such that, for example, the experiences of black women predicted by intersectional theory cannot be measured by statistical interactions (Dubrow 2008).

Given these difficulties, many intersectional theorists favor qualitative methods over quantitative methods for exploring intersectional issues (Stewart and McDermott 2004; Bowleg 2008). Furthermore, much of the preference for qualitative analysis seems to be due to an emphasis on studying the "context of lived experience" of members of some marginalized group (Jordan-Zachery 2007, 261). This seems to be better reflected by qualitative, as opposed to quantitative, work. However, qualitative analyses often involve one-on-one interviews or observations. Because these interviews take a significant amount of time for each participant, these qualitative studies often focus on individuals who are thought to belong (perhaps on the basis of a priori theory or background commitments) to marginalized groups. These methods are sensitive to the fallibility of the theories or background commitments that guide sample selection; researchers may miss out on the experiences of individuals who are not known ahead of time to be marginalized. Also, such a focus may promote the idea of marginalized groups as "Other" (Christensen and Jensen 2012). Further, emphasizing qualitative methods at the expense of quantitative methods may lead researchers to forfeit large survey data sets, often ones collected by a government agency or even cross-national ones, which can be very informative (Dubrow 2008).

We believe that the disagreement in the social scientific literature calls for pluralism about methods for testing intersectional hypotheses. To this end, we are responding to the call of quantitative researchers to develop better methods for testing intersectional claims with available large data sets. We see that our project is largely consistent with projects to improve qualitative methods by, for example, increasing sample sizes or extending the interviews to nonmarginalized groups. Our article will provide a method of translating certain characteristic claims of the intersectionality literature into statistically testable conjectures, which we hope will be of use to both social scientists and intersectional theorists elsewhere in the academy. In particular, we argue that interventionism and causal graphical modeling can provide a framework for testing claims about nonadditive effects and qualitative shifts in causal structure based on the intersection of certain variables. Note that we will not claim that these conjectures represent the entirety of intersectional analysis; our stance is compatible with the pluralism that acknowledges that the variety of problems addressed by means of intersectional analysis calls for a variety of methodologies.

2. Intersectionality Theory. Our analysis will address social scientists' concerns about applying intersectionality analysis by providing causal explications of some key claims in the intersectionality literature. To explicate a concept is to make it more precise in order to resolve ambiguities in such a way that the concept can be fruitfully applied in future research (Leitgeb 2013, 271). There is a demand in the social scientific literature for clarification of the methods by which claims made within intersectionality analysis can be tested. We believe that our explication of key intersectional concepts—which we shall call 'switch intersectionality' and 'nonadditive intersectionality'—will meet that demand by providing precise conditions under which intersectionality claims can be (dis)confirmed. It is this increase in the ability to draw out predictions that we believe vindicates our explication of intersectional concepts. We also believe that our explication is well grounded in the literature that has appeared on intersectionality. In this section we detail the relationship between our explication and previous work on intersectional theory.

Consider the following claim: "The first core idea of intersectional knowledge projects stresses that systems of power . . . cannot be understood in isolation from one another; instead, systems of power intersect and coproduce one another to result in unequal material realities and the distinctive social experiences that characterise them" (Collins and Chepp 2013, 60). This quote is taken from an article on intersectionality in the *Oxford Handbook of Gender and Politics* and is thus a deliberate attempt at stating the key principles of intersectionality studies for as wide an audience as pos-

sible. The claim admits of multiple interpretations, and we wish to draw attention to one in particular: the causal interpretation. Under this interpretation, when it is said that systems of power "intersect and coproduce one another," we interpret it as meaning that when given systems of power intersect, they produce causal effects on individuals (or groups) that they would not produce, or would not produce in the same way, if they did not intersect. We will explore two specific ways the causal interpretation of intersectionality analysis can manifest itself in the social sciences.

The first claim from within intersectional theory we explicate is what we call 'nonadditive intersectionality', which is the claim that somebody's intersectional identity can influence his or her life more than one would realize if one considered each of the identity categories separately (Weldon 2006; Hancock 2007; Bowleg 2008). We interpret this as meaning that some causal effects of belonging to multiple identity categories are stronger than one might have predicted from information about the causal effect of belonging to each identity category considered separately. Take, for instance, the claim made here: "In some cases the negative effects of racism and sexism might multiply each other, rendering women of color most disadvantaged on a dependent variable (e.g., income)" (Cole 2009, 177). There is already a causal effect of being a woman on one's income, and likewise there is already a causal effect of being a person of color. The intersectional phenomenon Cole reports is that occupying the intersectional identity of being a woman of color serves to amplify these causal processes.

The second claim to be explicated is what we call 'switch intersectionality'. Such claims describe causal relationships that are (de)activated only for individuals who occupy the intersection of certain identity positions. Consider, for instance, the following point from Dotson (2014, 52): "[There exists a] tendency to theoretically erase the experiences of oppression that are invoked as a result of being black women and not merely being black or a woman." We believe it is consistent with the author's intentions to say that combating this tendency involves acknowledging that the fact that a person is a black woman, rather than black or a woman considered singularly, causes her to undergo certain experiences. We will provide an analysis of switch intersectionality, which is to say causal processes that are activated only when the individuals under study occupy particular intersections of demographic categories.

In addition to providing a framework for stating and testing claims about switch or nonadditive intersectionality in social research, we also hope to contribute to ongoing debates in intersectionality theory itself. As noted in section 1, there has been a debate over the comparative virtues of qualitative and quantitative studies on intersectionality. We will offer a partial defense of quantitative studies of intersectionality. In particular, we will argue that

the following two claims are false. The first claim is that quantitative methods cannot detect the complexities with which intersectional theory is concerned: "Standard social scientific methodological techniques that attempt to isolate the effects of gender by controlling for race/ethnicity, or to isolate the effects of race/ethnicity by controlling for gender, are at odds with any attempt to study the complex interaction of race-gender in an organization. . . . Quantitative techniques designed to reveal uniformities of behavior are by design insensitive to difference, treating anything that deviates from the norm as an outlier or anomaly" (Hawkesworth 2006, 216–17; see also Carastathis 2014, 308). Switch intersectionality seems to be a prime example of intersectional analysis focusing on different effects introduced by the complexities of interactions between demographic categories. In particular, it is designed to aid the quantitative study of systematic qualitative differences in causal effects that arise from demographic differences within the population. The second claim we hope to show is false comes from Bowleg (2008). It is claimed there that quantitative methods are simply unable to make sense of the nonadditive intersectionality: "Alas, what holds in theory does not always translate easily to practice. Indeed, I would argue that it is virtually impossible, particularly in quantitative research, to ask questions about intersectionality that are not inherently additive" (314). But our analysis of nonadditive intersectionality aims to do just that. Hence, if our explication succeeds, it will show that quantitative techniques are adequate for stating and testing claims concerning nonadditive intersectional effects.

Some work in the quantitative social sciences claims to capture intersectionality by using traditional regression techniques along with interaction terms (e.g., Hinze, Lin, and Andersson 2012). Interaction terms are included in a regression model to account for the possibility that the best predictive model will be different for specific subpopulations in the sample. But these modeling techniques are purely predictive, not causal. That is, a statistically significant interaction term does not necessarily indicate how or if the interacting categories cause the outcome of interest; it tells us nothing about how the outcome might change if the system were manipulated in some way. Intersectionality theorists regularly use causal language in their descriptions of the phenomena, and they are often interested in guiding policy to produce better outcomes—not only in making predictions about future data points from present ones. If this is right, then they ought to be interested in estimating causal parameters, not just correlations. Regression analysis can be used to estimate causal parameters only under certain conditions on the data (either randomized treatment or control for all possible confounding, etc.). Without data that satisfy these conditions, regression estimates are measures of only association, not causation. To derive causal conclusions from purely associational results is to commit a well-known fallacy.

Finally, our framework could help address criticisms that have been leveled at intersectionality theory. Perhaps unsurprisingly in light of the disarray documented in the social science literature, intersectionality theorists have been criticized for not implicitly or explicitly outlining a methodology (Mutua 2013, 364). As we shall demonstrate in the next section, the causal interpretation of intersectionality theory comes prepackaged with an attendant methodology and so avoids this line of attack. More substantially, proponents of postintersectional and multidimensionality analysis have charged intersectionality theorists with making false predictions about, for instance, the social consequences of being perceived to be a heterosexual African American male in the twentieth-century United States (Mutua 2013, 358). However, as Mutua shows, it is not clear that intersectionality theorists were actually committed to the erroneous predictions these critics attributed to them. To explain why critics thought these phenomena were inconsistent with intersectionality theory, Mutua notes that intersectionality theorists typically concentrate on some demographics more than others, and hence it was not clear how to expand their ideas beyond these groups. Mutua advocates providing 'thick descriptions', which are in-depth qualitative accounts, of a greater variety of persons as the proper method of carrying out further intersectional studies to avoid running into similar criticisms in future (364).

Such qualitative work is important and useful on other grounds, but we do not believe it will address the problem Mutua has identified. Thick descriptions are by their nature particularist, and they will always be difficult to generalize from. While we anticipate that many intersectionality theorists would welcome a theory that resists supporting generalizations, we do not think that should be celebrated in this case. For, as Mutua's analysis exemplifies, in this case the inability to support generalizations makes intersectionality theory much harder to falsify and thus risks rendering the theory devoid of empirical content (Popper 2008, 45–48; Curry, forthcoming). The causal interpretation of intersectionality theory does not commit the intersectionality theorist to any specific substantial generalization about social life. It does, however, make it clear what sort of things should be tested for if one wishes to search for certain sorts of intersectional effects within any given population. To that extent it supports generalizations of intersectionality theory and renders it falsifiable.

Note that while we hope to support intersectionality theory by producing our explication, in this piece we will not directly vindicate the methods or claims of intersectionality theory by demonstrating their successful application on some data set. It remains open to the critic of intersectionality theory to argue that there are no interesting causal claims of the sort we explicate that have been or will be confirmed by the intersectionality theorist. Nonetheless, the application of these methods should advance the dialectic between in-

tersectionality theorists and their critics. If intersectionality theorists apply these methods and generate confirmed interesting causal claims about the structure of social life, it will represent a significant vindication of their theory. If, on the other hand, it turns out that when these methods are applied one still cannot confirm any interesting causal claims about the structure of social life, this would strengthen critics' case. Neither the original qualitative methods first proposed nor our new causal statistical machinery would have been able to generate evidence for the general claims of intersectionality theorists. This would suggest, against both Mutua's claims and our own beliefs, that the apparently failed predictions of intersectionality theory in the cases discussed by postintersectionality theorists were not the result of a particular flawed methodology and are perhaps instead indicative of a deeper flaw in the theory. Hence, whether the application of our proposed methods should result in well-confirmed causal claims for intersectionality theory or not, the methods outlined here should be of interest to all those interested in the empirical adequacy of intersectionality theory.

The causal interpretation of intersectionality thus serves three purposes. First, it provides a plausible and precise interpretation of some of the characteristic claims of intersectional theory. It therefore sheds light on the content of previously established theory. Second, as we shall show in the following sections, it allows us to give empirical content to intersectional claims in terms that are familiar to statistical social science. Hence it allows for the theory to be both applied with greater ease and submitted to more rigorous testing. Third, it addresses concerns authors have had about intersectionality theory and the role of quantitative methods in the study of intersectionality. We use the machinery of causal graphical modeling to carry out the explication that constitutes the causal interpretation of intersectionality and fulfill these purposes. That machinery is described in the following section.

3. A Brief Introduction to Causal Graphical Modeling. Causal graphical models, and causal Bayesian networks (CBNs) in particular, are representational tools for making causal inferences from data. The following presentation will be broad and mostly informal.[1] Here we will highlight facts relevant to just two potential applications of causal graphical models that might be useful to intersectional social theorists: representing causal hypotheses and searching for them from data algorithmically. Note that the meaning of "cause" is deliberately left unspecified here; the formal mod-

1. See Spirtes, Glymour, and Scheines (2000) or Pearl (2009) for book-length introductions to the topic; Greenland, Pearl, and Robins (1999) is an article-length exposition aimed at social scientists (epidemiologists).

eling and associated inference rules can be useful for a number of different understandings of causation. CBNs can aid in making inferences regarding counterfactuals, mechanisms, or the outcomes of interventions. In social science applications, the outcomes of interventions are often of particular interest, and so we will use interventionist language in the present discussion. We raise some issues with the interventionist interpretation in this context in section 5.

We begin with some terminological preliminaries. A graph consists of a set of vertices (or nodes) that are random variables, connected by edges. The random variables can be either continuous or categorical (i.e., they can take on a finite set of discrete values). Edges represent direct causal connections between the variables: if $X \rightarrow Y$, then X is a cause of Y. We also say that X is a parent of Y and Y is a child of X. A sequence of edges is a path. If there is a path from X to Z that consists of directed edges with arrowheads toward Z (a directed path), then X is an ancestor of Z and Z is a descendant of X. To make matters concrete, consider the graph in figure 1. This is a directed acyclic graph (DAG), meaning that all the edges are directed (one arrowhead, one direction) and that there are no cycles (no directed paths that start and end at the same vertex).

Assume that all the variables are categorical: the values of Parental Income, Education, Wealth, Gender, and Race are each discretized into, for example, two to four categories. In this graph Parental Income and Education are both direct causes of Wealth, and additionally, Parental Income is an indirect cause of Wealth via its influence on Education. Suppose we know the joint probability distribution over all our measured variables for some population. It is standard in the literature to make at least two assumptions about the relationship between our causal model and the joint probability distribution. These assumptions refer to probabilistic independence. Two variables X and Y are (unconditionally) independent if their joint probability distribution factorizes, that is, if $P(X, Y) = P(X)P(Y)$. Informally, this means that learning the value of Y yields no information about X. The variables X and Y are conditionally independent given Z if

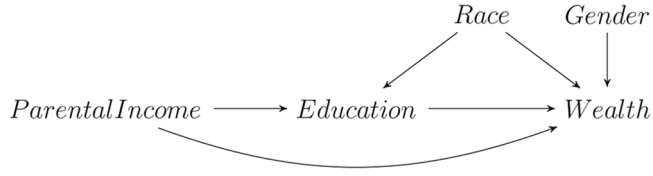


Figure 1. Directed acyclic graph (DAG).

https://doi.org/10.1086/684173 Published online by Cambridge University Press

$P(X, Y|Z) = P(X|Z)P(Y|Z)$. Informally, this means that learning the value of Y when the value of Z is already known yields no information about X. The first assumption is the causal Markov condition, which says that the joint probability distribution for the variables in the graph factorizes in a certain way: specifically, every variable in the graph is conditionally independent of its nondescendants, given its parents. This means that Education is independent of Gender given Race and Parental Income. The second assumption is called faithfulness, which requires that the only independencies reflected in the graphical structure are the independencies entailed by the causal Markov condition. This is a way of saying there is no "accidental canceling out" of causal pathways, making variables appear to be independent when they are in fact causally connected (and therefore dependent). See Spirtes et al. (2000, 29–42) for discussion of these assumptions. These two basic assumptions are ubiquitous in causal modeling, at least in social science applications.

The absence of an edge between two variables indicates that they are probabilistically independent given some subset of the other variables in the graph, including possibly the empty set. Race and Parental Income are independent in figure 1. However, the graph also predicts that Race and Parental Income are dependent conditional on Education. The reason is that the three variables form an unshielded collider at Education. A collider is a triple $X \rightarrow Y \leftarrow Z$ that has arrowheads "colliding" at one of the variables. A collider is unshielded if it has no edge between X and Z (the "colliding" variables). In an unshielded collider like Race → Education ← Parental Income, the colliding variables are independent unconditionally but conditionally dependent given Education. This is just one example of how graphical structure in a DAG relates to conditional independence facts in data. When social theorists use CBNs to represent a particular hypothesis, they should keep in mind such correspondences between graphical structure and independence facts. In particular, only if the social theorist believes that Education is in fact independent of Gender given Race (i.e., learning a person's gender is irrelevant for predicting his or her educational attainment when race is held fixed) does the structure in figure 1 accurately model the social system under investigation.

Such connections between graphical structure and probabilistic independence can be exploited in order to search for graphical models from data: the same kind of observational data that are typically available in quantitative sociological data sets (or in epidemiology, political science, or macroeconomics). Different algorithms are appropriate for different kinds of data. An overview of these algorithms would be inappropriate here; but see Eberhardt (2009) and the references therein for a more comprehensive introduction to the variety of search procedures and their requisite assump-

tions. The key idea is that researchers can use the same kind of observational data collected by government agencies, universities, hospitals, or other sources to learn causal structure.[2]

Once the causal structure is known—whether by application of search algorithms, domain-specific background knowledge, or other means—the researcher may quantitatively estimate causal effects of interest. Typically in the social sciences, causal effects are associated with the outcomes of interventions. The probability distribution of Y when X is set to some specific value x can be represented using Pearl's notation: $P(Y|do(X = x))$. The expression $do(X = x)$ is a way of saying that X is forced to take the value x, by some policy implementation or controlled experiment. This is generally distinct from $P(Y|X = x)$, which represents the distribution of Y in the subpopulation for which $X = x$. To illustrate this difference, think of the distribution of wealth in the subpopulation of Americans who have a bachelor's degree, that is, $P(\text{Wealth}|\text{Education} = \text{Bachelor's})$, as distinct from the distribution of wealth when all Americans are somehow forced to get a bachelor's degree, that is, $P(\text{Wealth}|do(\text{Education} = \text{Bachelor's}))$.[3] Explicating which interventionist quantity is referred to by "causal effect" depends on context, in particular, what kind of data the researcher is working with. For example, when working with linear and continuous variables, it is common to define the total causal effect of X on Y as $(\partial/\partial x)\mathbf{E}(Y \mid do(X = x))|_{x=x'}$. This is the rate of change in the expectation of Y as X is forced to vary. When working with binary variables, one might define the total causal effect of X on Y as $\mathbf{E}[Y|do(X = 1)] - \mathbf{E}[Y|do(X = 0)]$. If the purported cause variable can take on more than two values, the causal effect is usually defined with respect to some reference value. For example, the causal effects of $do(\text{Race} = \text{White})$ and $do(\text{Race} = \text{Asian})$ on W (wealth) can be defined with Race = Black as a reference value: $\mathbf{E}[W|do(\text{Race} = \text{White})] - \mathbf{E}[W|do(\text{Race} = \text{Black})]$ and $\mathbf{E}[W|do(\text{Race} = \text{Asian})] - \mathbf{E}[W|do(\text{Race} = \text{Black})]$, respectively. The choice of reference value is important and usually reflects something about the researcher's question of interest. Note that total effects are distinct from direct effects, where the former is a sort of combination of all the causal pathways between X and Y and the latter is only a measure of the direct connection.

2. For constraint-based search procedures such as PC and FCI, see Spirtes et al. (2000). For a Bayesian score-based algorithm, see Chickering and Meek (2002). For linear structural equation models with non-Gaussian noise, see Shimizu et al. (2006) and Hoyer et al. (2008). These are just a few examples; the number and variety of algorithms for causal search have exploded in recent years.

3. See Meek and Glymour (1994) for discussion of the difference between conditioning and intervening.

There are numerous other "causal effect" quantities that a researcher might be interested in.[4] The key point is that to accurately estimate these quantities from observational data, it is necessary to know something about the causal structure; they are not generally derivable from correlations alone. But once the causal structure is known, there are a number of techniques for obtaining consistent numerical estimates of these quantities from data.

We should mention that there are alternative methods for estimating many of these quantities of interest that are not graphical, for example, Rubin's (2005) potential outcomes approach to causality. Rubin's framework is mathematically equivalent to the CBN approach, but we think that the graphical approach has some advantages. In particular, the graphical representation is useful for tractable and careful representation of causal hypotheses in domains with a large number of variables, it facilitates reliable and fast inference from observational data, and this framework comes equipped with methods for algorithmic search. There has been recent work directly relevant to the topic of this article within the potential outcomes framework (Egami and Imai 2015; VanderWeele 2015). The authors define a quantity called average treatment interaction effect, which has two interpretations. These interpretations correspond quite nicely to what we refer to as 'nonadditive intersectionality' and 'switch intersectionality'. An important difference, however, is that in this alternative formalism the two interpretations are mathematically equivalent, whereas in our discussion the two ideas can be formally distinct; they may be equivalent under certain model parameterizations, but our present discussion is fully general and nonparametric.

4. Causal Interpretations of Intersectionality. In this section we offer some interpretations of intersectionality claims by making use of the causal modeling framework. We begin with nonadditive intersectionality. One claim made by Cole in section 2 is that the social or economic effects of being a woman of color are not simply the sum of the effects of being a person of color and being a woman. We take this to mean that the magnitude of the causal effect (on some dependent variable) of being a woman of color is not equal to the sum of the magnitudes of the causal effects of being a person of color and being a woman. Such a claim relies on the intelligibility of measuring the "strength" or "magnitude" of causal effects. How to spell this out more precisely depends on the interpretation of causation intended. Often, researchers have in mind the kind of interventionist quantities just discussed in the previous section. In such cases there is a straightforward way to represent the above claim. Let G represent gender ($0 =$ male, $1 =$ female) and R represent race ($0 =$ white, $1 =$ black). Both variables are being treated

4. See Pearl (2001) for definitions and discussion.

as binary only for simplicity here. The causal effect of gender on some dependent variable, for example, W (wealth), can be written

$$\theta_G := \mathbf{E}[W \mid do(G = 1)] - \mathbf{E}[W \mid do(G = 0)]. \quad (1)$$

That is the difference (with respect to wealth) between intervening to make all members of the population perceived as female versus perceived as male. Similarly, the causal effect of race on W can be written

$$\theta_R := \mathbf{E}[W \mid do(R = 1)] - \mathbf{E}[W \mid do(R = 0)], \quad (2)$$

and the expected difference associated with being perceived as white male versus black female is

$$\theta_{GR} := \mathbf{E}[W \mid do(G = 1 \text{ and } R = 1)] - \mathbf{E}[W \mid do(G = 0 \text{ and } R = 0)]. \quad (3)$$

So one way of explicating an intersectionality hypothesis with respect to gender, race, and wealth is that the effects are nonadditive:

$$\theta_{GR} \neq \theta_G + \theta_R. \quad (4)$$

The intersectional theorist might also hypothesize that θ_{GR} is some specific function of θ_G and θ_R (perhaps it is the multiplicative product, as the Cole excerpt in sec. 2 suggests). In another context, the "causal effect" of interest might not be the total causal effect, but it might be the direct effect, or the causal effect when certain other variables in the system are held fixed. Intersectionality theorists have also made claims about the intersection of other social categories (e.g., age, citizenship status, or sexual orientation), and analogous quantitative statements can be formulated for such hypotheses. With reliable estimates of the causal quantities involved, such hypotheses can be represented and tested against data. The framework allows for representations of privilege on par with representations of oppression or exclusion; being a white woman might involve certain advantages in some contexts (relative to other demographic categories under consideration), and such claims can be straightforwardly expressed.

Switch intersectionality claims indicate that some categories interact to produce novel effects, that is, that there are effects associated with occupying the intersection of multiple categories that are not present for individuals 'outside' the intersection (individuals who occupy only some of the categories but not all of them). One of Crenshaw's paradigmatic examples relates barriers to access to social resources to an individual's gender and immigration status (1991, 1245–46). Immigrant women, and in particular immigrant women of color, face unique obstacles to accessing certain social resources (e.g., aid from domestic violence shelters). These are obstacles that nonimmigrant white women (perhaps because they are more likely to

https://doi.org/10.1086/684173 Published online by Cambridge University Press

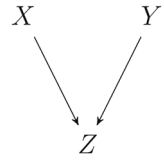
speak fluent English) are less likely to face; (non-)immigrant men, too, are not likely to encounter these particular obstacles because they are less likely to require the services of domestic violence shelters, in Crenshaw's example. So, occupying the intersection of certain categories—woman, immigrant, person of color—is associated with a specific effect. When demographic categories and effects are represented by random variables, this phenomenon can be understood as the emergence of effects that are active only when some variables take on specific values. That is, only when Gender, Immigration Status, and Race take on particular values is Resource Access 'switched on' as an effect.[5]

Unfortunately, the standard framework of CBNs is not (by itself) very useful for representing such situations. By definition, Z is dependent on X if there is any value of X that changes the probability distribution over Z. And since causal connections in CBNs are grounded in dependence facts, graphical representations are insensitive to 'context-specific' details. For example, if X is a cause of Z only when $Y = 1$ (but not when $Y = 0$), we would still represent the system as in figure 2*a*. As we can see from the conditional probability table in figure 2*b*, only when $Y = 1$ does changing the value of X affect the probability of $Z = z$. When $Y = 0$, X and Z are independent. Spirtes et al. (2000, 24–25) discuss this limitation of the CBN framework, and they point out that a graph like the one in figure 2*a* is correct but not fully informative; there is information in figure 2*b* that is masked by the "global" definition of independence. Incorporating context-specific ("local") independence of this sort in graphical modeling is an active area of research, especially in computer science. Several augmentations to the Bayesian network have been proposed and explored (Geiger and Heckerman 1991; Boutilier et al. 1996; Chickering, Heckerman, and Meek 1997; Friedman and Goldszmidt 1998). Work in this area incorporates facts about independences that hold only in certain contexts, that is, only when certain variables take on particular values. Most simply, we can represent such information with conditional independence tables as in figure 2*b*, but more elaborate supplements or changes to Bayesian network representations such as multinets, similarity networks, decision trees, and decision graphs have proved fruitful in the design of certain algorithms. We will briefly illustrate the multinet representation, following Geiger and Heckerman (1991), since it is most similar to the Bayesian network representation already introduced.

A multinet is just a set of Bayesian networks, where each graph in the set is "localized" to a specific variable assignment for one variable. Suppose we have collected data on three variables: Immigration Status, Incarceration

5. Note that a causal connection can have positive or negative strength: so causes can both 'promote' and 'inhibit' their effects.

a



b

X	Y	$P(Z\|X,Y)$
0	0	p_1
1	0	p_1
0	1	p_2
1	1	p_3

Figure 2. *a*, Causal DAG; *b*, probability table.

Rate, and National Origin. The last variable is partitioned into three values: one's National Origin can be equal to 'United States', 'Latin America', or 'elsewhere'. In the southwestern United States, border patrol officers target individuals believed to be from Latin America, entering the country without documentation (Miller 2010). There are of course immigrants from elsewhere in the population, but they are not the targets of American border patrol and so are less likely to be incarcerated. Among individuals from Latin America then, Immigration Status is a cause of Incarceration Rate. Among individuals from the United States or elsewhere, perceived National Origin is a common cause of both Immigration Status (since individuals with origin in the United States are almost definitely citizens) and Incarceration Rate. Also, among individuals from the United States or elsewhere, learning their Immigration Status is not informative for predicting Incarceration Rate when the value of National Origin is known; Immigration Status and Incarceration Rate are made independent by conditioning on their common cause. The multinet representation of this hypothetical is in figure 3. This is one way of presenting the idea that immigrants from Latin America experience a novel effect. Note that we are departing from the usual causal semantics associated with CBNs. In a CBN, X causes Y if there is a directed edge or a sequence of directed edges from X to Y. Although there is only a directed edge between Immigration Status and Incarceration Rate in the second graph, the understanding is that immigration status is only causally efficacious among individuals from Latin America. There is no sense in which either National Origin or Immigration Status is "the" singular cause of Incarceration Rate: the two variables interact to affect Incarceration Rate.

There are a number of caveats to representing switch intersectionality with multinets. Care must be taken in interpreting multinets causally. In many computer science applications, the intended use of graphical models is efficient calculation of conditional probabilities, not the outcomes of interventions. So, the apparatus is not explicitly causal, and the causal inter-

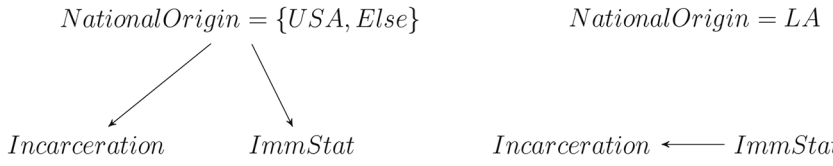


Figure 3. Multinet. Here we use Incarceration and ImmStat as shorthand for Incarceration Rate and Immigration Status. LA is shorthand for Latin America and Else is shorthand for elsewhere.

pretation of work in this area is currently underinvestigated. In particular, if the variable that is partitioned and conditioned on across graphs in a multinet (National Origin in our example) is an effect of other variables in the graph, new dependencies can be introduced by conditioning that are perhaps incorrect to interpret causally. If the variables in figure 3 were embedded in a larger causal structure that had an unshielded collider at National Origin, the colliding variables would become dependent when we condition on National Origin to create the individual graphs in the multinet. That dependence would introduce an edge between the colliding variables that should not be interpreted causally. See Geiger and Heckerman (1991, 121) for discussion. Multinets are thus more promising representations when the partitioned and conditioned variable (e.g., National Origin) has no causes; that is, it has no parents in the set of measured variables. Finally, it should be noted that search procedures for discovering multinet structure from data are less developed as compared with CBNs.

The considerations in this section indicate that intersectionality claims can be effectively investigated with quantitative methods. The appropriate representations and mathematical tools depend on the kind of claim being investigated and the kind of data available. We hope that the social researcher interested in incorporating intersectional analysis in her work can benefit from some of these tools.

5. Demographic Categories as Causal Variables. Before concluding, we anticipate and forestall some objections that theorists may have to our explication of intersectional concepts. For, in addition to the worries about quantitative methods in intersectionality theory referred to in section 1, there have been worries expressed about statistical measures of race as a cause in general (Bonilla-Silva and Zuberi 2008) and about the use of the CBN framework in particular to study the causal effects of membership in demographic categories (Holland 2008). Since readers may be familiar with such concerns, we address them here.

The general worry about the use of statistical measures of race as a cause is that they will tend to promote the reification of racial categories. Such

https://doi.org/10.1086/684173 Published online by Cambridge University Press

a reification has historically been associated with pernicious social movements and is in any event scientifically implausible (Bonilla-Silva and Zuberi 2008, 6; Maglo and Martin 2012). Whatever merits there are in this argument, we do not believe that it counts against the work we have done here. We are explicating claims made within the intersectionality literature. If we are correct, authors in that literature were already committed to the idea that membership within demographic categories plays a causal role in bringing about the life experiences that people undergo. The fact that it may not have been expressed in formal terminology does not seem relevant to whether this is objectionably reifying or not. Rather, it is the fact that demographic categories are playing any sort of causal role at all that drives this line of critique. We have not invented any such causal claims, but rather we have provided a causal modeling framework for their more explicit representation. If intersectionality theorists decide to revise or reject prior claims as a consequence of seeing them made fully explicit, then in one sense our task has been successful.

The more specific worry about the CBN framework concerns its link to the interventionist framework. Some social scientists claim that we cannot intervene on a person's gender or race, so we cannot properly treat these properties as causes of anything but rather only as proxies (e.g., Woodward 2003, 115; VanderWeele and Robinson 2014). If interventions are required to understand the empirical content of causal claims in the CBN framework, then this seems to count them out as potential causes. But we have claimed that an advantage of our theory is that it allows the application of quantitative methods to give empirical content to intersectionality theory, a branch of theory that requires the use of demographic categories. This objection thus gets at the heart of our project.

Our first reply is to note that the CBN framework is not inextricably bound to interventionism (Glymour and Glymour 2014). Hence even if one objects to, or rejects, this understanding of causation, one can still benefit from using CBNs to represent one's causal claims. In tying causal structure to a particular sort of graph structure, the CBN framework facilitates the full and explicit recognition of all the factors one thinks are relevant to the situation at hand. As we hope our diagrams show, this can encourage the clear representation of one's claims and thus make it easier for both theorists and critics to get a handle on whatever model is being used in a given analysis. Furthermore, if this alternative causal semantics satisfies some minimal conditions like, for example, 'causes change the probabilities of their effects', then facts about conditional independencies are still relevant and could significantly narrow down the space of empirically adequate models for a given data set.

We should note that in particular our explication of 'switch intersectionality' does not depend in any way on an interventionist semantics for

causation. However, 'nonadditive intersectionality' does as we described it. We are not aware of any way of (quantitatively) spelling out the comparison of 'strength' of causal links that does not assume some kind of interventionist account of causation. Perhaps there is some alternative measure of causal strength between two variables that does not refer to interventions; in this case the researcher can adopt an analogous scheme for talking about when the strengths of multiple causal connections combine in a particular functional way. So one can view our explication of nonadditive intersectionality as a general recipe for hypotheses that are both causal and nonadditive, even though our own mathematical representation is wedded to interventionism.

However, our second and more substantive reply is that intersectionality theorists should adopt an interventionist framework. We consider it an advantage of interventionist interpretations of causation that they require those inclined to make causal claims to be explicit about what exactly would have to change in order to bring about a difference. Woodward argues a similar point and uses the example of the (apparent) effects of gender on hiring decisions to illustrate the idea (2003, 114–17). Being explicit about the interventions one believes necessary to make a difference on hiring decisions can cause one to realize that an applicant's gender is not the cause of her hiring outcome. Often, perceived gender and perceived race can be manipulated to change hiring outcomes, and we see that in some experimental work on hiring bias, perceived race and perceived gender are experimentally controlled (e.g., Moss-Racusin et al. 2012). In these cases, the true cause of the hiring decision is something about the wider patriarchal social structure in which the hiring takes place, in the sense that manipulating the social structure while keeping the applicant's gender fixed would lead to different hiring outcomes. Woodward's point seems to mirror intersectionality theorists' talk of interlocking 'systems of power' or 'systems of oppression' bringing about consequences for differently located individuals, some of which we quoted in section 2. This suggests that macro-level social facts—for instance, that there exist widespread discriminatory employment practices—are the causes of interest to intersectionality theorists.

Note, however, that Woodward's interventionist framework may not be fully adequate to capture all the macro-level causes that social theorists may attribute to particular outcomes. Woodward places restrictions on what is the right kind of intervention (2003, 98). It is an open question of broad social concern whether interventions on macro-level phenomena such as 'the patriarchy' or 'global capitalism' could ever satisfy the conditions Woodward outlines. If not, then in Woodward's theory these phenomena cannot be causes at all. Perhaps interventionist talk needs some further refinement to capture these phenomena of interest. In particular, it may be usefully refined by being built into a theory of social structural explanation,

as outlined by Haslanger (2015). To give a Haslangian social structural explanation of some phenomenon is to outline a system of relations people (or animals or objects of any sort) stand in to each other then show that occupants of certain positions in this system are bound by whatever constraints the system induces to act in ways that produce the phenomenon. For instance, Haslanger discusses a social structural explanation of the differences in average level of economic power of men and women by noting that a couple consisting of a man and a woman who live in a capitalistic society without affordable child care and in which there is a wage gap between men and women will, within such a structure, tend to act rationally by having the woman drop out of the labor force upon the birth of a child and thus accrue greater economic power to the man (Haslanger 2015, 10). Nothing particular about the man or woman involved was necessary to state this explanation; we simply needed to describe some facts about the social structure they exist within and note the constraints these facts place on the couple's decision making. We believe that in order to make sense of the claims about macrosocial causation that intersectionality theorists seem to rely on within an interventionist framework, it would be useful to develop a CBN interpretation of Haslangian social structural explanation wherein the organization of the graph represents the social structure in question. We hope to explore this idea in future research.

These considerations serve to illustrate the general moral: causal claims that intersectionality theorists are interested in and that are based on demographic categories are often ambiguous in an important sense, and being explicit about the relevant interventions can serve to "disambiguate" them. Hence, despite the apparent tension, the interventionist framework may cohere especially well with the concerns of intersectionality theorists by pressing scholars to pay explicit attention to the sort of proxies that might be used for demographic categories when making causal claims. We hope that our explication actually prompts future research on intersectional theory to investigate the question of what role (if any) demographic categories should play in causal explanations of social phenomena.

Consideration of these objections thus actually served a dual purpose. First, it is important that social scientists be aware of these worries when they apply the CBN framework. A common feature of all our replies to these lines of critique is that the only way of avoiding these problems is to explicitly consider them as one works. But, second, they highlight a further benefit to the causal interpretation of intersectionality analysis. Previous theoretical work was, we claim, already bound up in particular causal claims. However, since people were not explicit about the causal status of their claims, these issues have not yet been addressed in the context of intersectionality theory. Simply by being explicit, the causal interpretation of intersectionality encourages conscious consideration of these issues by

intersectionality theorists. This increase in self-awareness is itself a considerable advantage, especially since it draws attention to the question of what interventions would be necessary to bring about relevant events.

6. Conclusion. We have responded to social scientists' call for concrete predictions and testable hypotheses from intersectional theory and provided social scientists one way of quantitatively testing these hypotheses. At the same time, we have tried to stay true to some of the prominent claims of intersectional theory: switch intersectionality and nonadditive intersectionality. The former states that systems of power intersect such that they produce effects on an individual that they would not produce (in the same way) if the systems of power did not intersect; the latter states that individuals' intersectional identity can influence their life more than one would expect by merely "adding" the effects on each group of which they are a member. Both feature prominently in descriptions of intersectional theory by intersectional theorists.

Causal modeling lends itself particularly well to testing for these two concepts of intersectional theory in a given sample. CBNs in conjunction with a multinet augmentation are just one type of tool for social scientists to examine intersectional hypotheses. However, when researchers adopt the interventionist framework, they gain a vocabulary with which to render explicit claims about potential interventions on perceived race and gender as well as to empirically describe the "interlocking systems of power" emphasized by intersectional theorists.

Of course, quantitative analysis has its difficulties, and the approach suggested here suffers from all the usual practical hurdles, for example, the availability of good, representative data sets. Also, when the "main effects" are strong, statistically significant interactions may be difficult to detect (Bowleg 2008). Nothing we propose alleviates these difficulties, but we hope to have illuminated a more sound methodology for those quantitative social scientists who are interested in exploring intersectional hypotheses.

The use of CBNs and particular augmentations can be immediately applied to current intersectional research questions. For example, Thompson et al. (forthcoming) examine potential factors in the early drop-off after introductory courses (and, thus, the underrepresentation) of both women and black students in philosophy. However, they have been criticized for failing to examine intersectional hypotheses about black women in philosophy (Freeman 2014). Although no explicit intersectional hypotheses have yet been proposed for the underrepresentation of black women in philosophy, the methods described in this article could be used in further research to explore whether the experiences in philosophy classrooms of black women are different compared to both black men and white women either in kind or in intensity.

Our analysis may also be used to test Antony's (2012) perfect storm hypothesis. This hypothesis suggests that there is no single factor that can explain why women are still severely underrepresented in philosophy compared to other disciplines such as biology and English, but rather a number of familiar factors "take on particular forms and force as they converge within the academic institution of philosophy" (231). Antony's perfect storm hypothesis is explicitly an intersectional hypothesis. These familiar factors may include phenomena such as stereotype threat, implicit bias, an emphasis on particular individuals in an entrenched canon, and the stereotype of a lone thinker. None of these factors by themselves is enough to explain why philosophy's demographic diversity has lagged behind many other fields; together these factors create a 'perfect storm'. This is a case of switch intersectionality. However, it is worth noting that the intersecting variables according to the perfect storm hypothesis include stereotype threat and diversity of authors on the syllabus rather than demographic variables. Although the factors that are interacting are not demographic variables in this case, we believe that the analysis presented in this article suggests the type of comprehensive data set that would be required to test the perfect storm hypothesis. While there may not be a data set combining evidence about implicit bias, stereotype threat, and other factors contributing to the underrepresentation of certain groups in philosophy, our analysis also provides a way to analyze the data for intersectional hypotheses were the data to be available.

Ultimately, we see this article as just one step in the right direction. Further research should investigate quantitative methods for testing hypotheses involving other features of intersectional theory. Research relating qualitative and quantitative methods would also be extremely fruitful for intersectional theorists and social scientists alike. Finally, by accepting the empirically testable interpretations of intersectional theory described here, intersectional theorists could provide new, concrete, and testable intersectional hypotheses.

REFERENCES

Antony, Louise. 2012. "Different Voices or Perfect Storm: Why Are There So Few Women in Philosophy?" *Journal of Social Philosophy* 43 (3): 227–55.

Bonilla-Silva, Eduardo, and Tukufu Zuberi. 2008. "Toward a Definition of White Logic and White Methods." In *White Logic, White Methods: Racism and Methodology*, ed. Tukufu Zuberi and Eduardo Bonilla-Silva. Lanham, MD: Rowman & Littlefield.

Boutilier, Craig, Nir Friedman, Moises Goldszmidt, and Daphne Koller. 1996. "Context-Specific Independence in Bayesian Networks." In *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*, 115–23. Burlington, MA: Morgan Kaufmann.

Bowleg, Lisa. 2008. "When Black + Lesbian + Woman ≠ Black Lesbian Woman: The Methodological Challenges of Qualitative and Quantitative Intersectionality Research." *Sex Roles* 59 (5–6): 312–25.

Carastathis, Aanna. 2014. "The Concept of Intersectionality in Feminist Theory." *Philosophy Compass* 9 (5): 304–14.

Chickering, David M., David Heckerman, and Christopher Meek. 1997. "A Bayesian Approach to Learning Bayesian Networks with Local Structure." In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, 80–89. Burlington, MA: Morgan Kaufmann.

Chickering, David M., and Christopher Meek. 2002. "Finding Optimal Bayesian Networks." In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, 94–102. Burlington, MA: Morgan Kaufmann.

Christensen, Anne-Dorte, and Sune Q. Jensen. 2012. "Doing Intersectional Analysis: Methodological Implications for Qualitative Research." *Nordic Journal of Feminist and Gender Research* 20 (2): 109–25.

Cole, Elizabeth. 2009. "Intersectionality and Research in Psychology." *American Psychologist* 64 (3): 170–80.

Collins, Patricia H., and Valerie Chepp. 2013. "Intersectionality." In *The Oxford Handbook of Gender and Politics*, ed. G. Wayla, K. Celis, J. Kantoha, and S. Weldon. Oxford: Oxford University Press.

Crenshaw, Kimberle. 1991. "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color." *Stanford Law Review* 43 (6): 1241–99.

Curry, Tommy J. Forthcoming. "Black Studies, Not Morality: Anti-black Racism, Neo-liberal Cooptation, and the Challenges to Black Studies under Intersectional Axioms." In *Emerging Voices of Africana: Disciplinary Resonances*, ed. M. Tillotson. Trenton, NJ: Third World–Red Sea.

Dotson, Kristie. 2014. "Making Sense: The Multistability of Oppression and the Importance of Intersectionality." In *Making Sense of Race and Gender: An Intersectional Approach*, ed. M. Goswami, M. M. O'Donovan, and L. Yount. London: Pickering & Chatto.

Dubrow, Joshua K. 2008. "How Can We Account for Intersectionality in Quantitative Analysis of Survey Data? Empirical Illustration for Central and Eastern Europe." *Research and Methods* 17 (5–6): 85–100.

Eberhardt, Frederick. 2009. "Introduction to the Epistemology of Causation." *Philosophy Compass* 4 (6): 931–45.

Egami, Naoki, and Kosuke Imai. 2015. "Causal Interaction in High-Dimension." Unpublished manuscript, Princeton University. http://imai.princeton.edu/research/files/int.pdf.

Freeman, Lauren. 2014. "Creating Safe Spaces: Strategies for Confronting Implicit and Explicit Bias and Stereotype Threat in the Classroom." *APA Newsletter on Feminism and Philosophy* 13 (2): 3–12.

Friedman, Nir, and Moises Goldszmidt. 1998. "Learning Bayesian Networks with Local Structure." In *Learning in Graphical Models*, ed. Michael I. Irwin, 421–59. Dordrecht: Springer.

Geiger, Dan, and David Heckerman. 1991 "Advances in Probabilistic Reasoning." In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, 118–26. Burlington, MA: Morgan Kaufmann.

Glymour, Clark N., and Madelyn R. Glymour. 2014. "Race and Sex Are Causes." *Epidemiology* 25 (4): 488–90.

Greenland, Sander, Judea Pearl, and James M. Robins. 1999. "Causal Diagrams for Epidemiologic Research." *Epidemiology* 10 (1): 37–48.

Hancock, Ange-Marie. 2007. "When Multiplication Doesn't Equal Quick Addition: Examining Intersectionality as a Research Paradigm." *Perspectives on Politics* 5 (1): 63–79.

Haslanger, Sally. 2015. "What Is (Social) Structural Explanation?" *Philosophical Studies* 31 (8116): 1–18.

Hawkesworth, Mary. 2006. *Feminist Inquiry: From Political Conviction to Methodological Innovation*. New Brunswick, NJ: Rutgers University Press.

Hinze, Susan W., Jieulu Lin, and Tanetta E. Andersson. 2012. "Can We Capture the Intersections? Older Black Women, Education, and Health." *Women's Health Issues* 22 (1): e91–e98.

Holland, Paul W. 2008. "Causation and Race." In *White Logic, White Methods: Racism and Methodology*, ed. Tukufu Zuberi and Eduardo Bonilla-Silva. Lanham, MD: Rowman & Littlefield.

Hoyer, Patrick O., Shohei Shimizu, Antti J. Kerminen, and Markus Palviainen. 2008. "Estimation of Causal Effects Using Linear Non-Gaussian Causal Models with Hidden Variables." *International Journal of Approximate Reasoning* 49 (2): 362–78.

Jordan-Zachery, Julia S. 2007. "Am I a Black Woman or a Woman Who Is Black? A Few Thoughts on the Meaning of Intersectionality." *Politics and Gender* 3 (2): 254–64.

Kantola, Johanna, and Kevät Nousiainen. 2009. "Institutionalizing Intersectionality in Europe: Introducing the Theme." *International Feminist Journal of Politics* 11 (4): 459–77.

Leitgeb, Hannes. 2013. "Scientific Philosophy, Mathematical Philosophy, and All That." *Metaphilosophy* 44 (3): 267–75.

Maglo, Koffi N., and Linda J. Martin. 2012. "Researching vs. Reifying Race: The Case of Obesity Research." *Humana Mente* 22 (1): 111–43.

McCall, Leslie. 2005. "The Complexity of Intersectionality." *Signs: Journal of Women in Culture and Society* 30 (3): 1771–1800.

Meek, Christopher, and Clark Glymour. 1994. "Conditioning and Intervening." *British Journal for the Philosophy of Science* 45 (4): 1001–21.

Miller, Todd. 2010. "Arizona, the Anti-immigrant Laboratory." NACLA Report on the Americas. https://nacla.org/node/6681.

Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." *Proceedings of the National Academy of Sciences* 109 (41): 16474–79.

Mutua, Athena D. 2013. "Multidimensionality Is to Masculinities as Intersectionality Is to Feminism." *Nevada Law Journal* 13 (2): 341–67.

Pearl, Judea. 2001. "Direct and Indirect Effects." In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. Burlington, MA: Morgan Kaufmann.

———. 2009. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press.

Popper, Karl. 2008. *Conjectures and Refutations*. New York: Routledge Classics.

Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100 (469): 322–31.

Shimizu, Shohei, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. 2006. "A Linear Non-Gaussian Acyclic Model for Causal Discovery." *Journal of Machine Learning Research* 7:2003–30.

Spirtes, Peter, Clark N. Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*. 2nd ed. Cambridge, MA: MIT Press.

Stewart, Abigail J., and Christa McDermott. 2004. "Gender in Psychology." *Annual Review of Psychology* 55:519–44.

Thompson, Morgan, Toni Adleberg, Sam Sims, and Eddy Nahmias. Forthcoming. "Why Do Women Leave Philosophy? Surveying Students at the Introductory Level." *Philosophers' Imprint*.

VanderWeele, Tyler J. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford: Oxford University Press.

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): 473–84.

Warner, Leah R. 2008. "A Best Practices Guide to Intersectional Approaches." *Sex Roles* 59 (5–6): 454–63.

Weldon, Laurel S. 2006. "The Structure of Intersectionality: A Comparative Politics of Gender." *Politics and Gender* 20 (6): 805–25.

Woodward, James. 2003. *Making Things Happen*. Oxford: Oxford University Press.