Received: 9 July 2022 | Revised: 25 June 2023 | Accepted: 27 June 2023

DOI: 10.1111/rode.13029

SPECIAL ISSUE ARTICLE

WILEY

Poor economics and its missing mechanisms: The case for causal mediation

Sunil Mitra Kumar[1] | Ragupathy Venkatachalam[2]

[1]King's India Institute and Department of International Development, King's College London, London, UK

[2]Institute of Management Studies, Goldsmiths, University of London, London, UK

Correspondence
Sunil Mitra Kumar, King's India Institute and Department of International Development, King's College London, London, UK.
Email: sunil.kumar@kcl.ac.uk

Abstract

A key aim of studying development is to understand the factors that shape socioeconomic progress and explain inequalities. In empirical work, the predominant focus has been on posing these questions in the language of causal inference: how one or more variables effect an outcome of interest, with the estimation of Average Treatment Effects (ATE) becoming prioritised as the key objective. The 'credibility revolution' and the emphasis on randomised controlled trials in research on development has cemented this dominance, because randomisation is well-suited to estimating the ATE. This paper argues that this dual dominance—ATE as main question of interest, and experiment as preferred method—is narrow and restrictive. We propose causal mediation frameworks as an alternative, which are routinely used in disciplines including epidemiology, psychology, sociology and political science where causal mechanisms are an equally important focus. We introduce key concepts and definitions of path-specific effects, and discuss identification and estimation approaches. We illustrate applications for development and demonstrate how causal mediation brings the focus back to contextual knowledge, combining this with empirical rigour.

KEYWORDS

causal mechanisms, causal mediation, development economics

1 | INTRODUCTION

Amongst the key aims of the study of development across disciplines is to understand the factors that shape socioeconomic progress and explain underlying, often persistent variations across groups and countries. Such factors can operate and be defined at macro level (e.g., nations, institutions, structural constraints) and micro level (e.g., individuals and households). In practice, interest in this larger query is often broken down into smaller sub-questions at different levels to make theoretical and empirical investigations tractable. These involve focusing on isolating a smaller subset of variables, developing a theoretical framework to explain observed empirical patterns, potential determinants, and where relevant offer policy insights. In development economics, the predominant focus has been on posing these questions in the language of causal inference: how one or more variables effect an outcome of interest. This interest in causal inference has shaped the field in distinct ways, with the estimation of Average Treatment Effects (ATE) becoming prioritised as the key objective, be it using experimental or observational data. For example, the effects of a job-training programme for the unemployed on future employment levels, or of a remedial teaching intervention on learning scores amongst school pupils and so on.

Experimental methods in particular have enjoyed disproportionate attention in the past two decades and there is a sizable literature on their advantages (e.g., Banerjee & Duflo, 2011; Haynes et al., 2012) and key limitations, both methodological and ethical (e.g., Deaton & Cartwright, 2018; Teele, 2014). The ATE focuses attention on the causal effect of a single variable on a single outcome. Randomisation directly enables this form of causal inference, while observational data can also be used to infer these effects using appropriate statistical adjustment.

However in our view, this dual dominance—of the ATE as main question of interest, and experiment as preferred method—represents a narrow and restrictive form of causal inference. The limitation is twofold. First, a narrow focus on what works has consequently led to neglecting causal mechanisms or why it works. The processes through which various factors together shape developmental outcomes are seldom the object of inquiry, or at best indirectly so. This is in contrast to disciplines like epidemiology, psychology, sociology and political science where causal mechanisms are an important focus. Second, a failure to fully exploit the power of causal inference frameworks to glean insights into such mechanisms from observational data has further narrowed the scope of scholarship.

In contrast, understanding development more holistically necessitates moving beyond measuring ATEs alone and the related disproportionate focus on experimental methods. Accounting for causal mechanisms is a crucial element of such understanding, and thereby the myriad ways in which context, structure and processes interact to shape developmental outcomes. In order to understand mechanisms, we need to identify and disentangle the pathways through which causal effects are manifested. This involves studying intermediate variables—mediators—and drawing upon contextual background knowledge to specify the causal structures involved. This also requires bringing the emphasis back to practitioner expertise and contextual or domain knowledge, which otherwise have a tendency to be relatively deprioritised.

In this paper, we illustrate this argument and offer a concrete methodological solution for investigating causal mechanisms through the toolbox of causal mediation methods. Mediation methods are common in a variety of disciplines including psychology, epidemiology, political science and sociology, so forth, but their adoption has been very limited within economics, and even less so within the study of development. For example, while Deaton (2010a, 2010b) has argued forcefully for the need to understand the mechanisms behind problems in development and to move beyond reduced-form approaches, he emphasises the role of structural equation modelling in pursuing these aims, but does not offer causal mediation methods as a way to go one step further. Imbens (2020) offers an authoritative comparison of graphical methods (which underpin causal mediation) and the potential outcome framework generally used in economics. While he alludes to the importance of mediation methods, his general hesitation towards graphical methods potentially undermines this message and might help explain their limited popularity in the discipline. Our attempt in this paper is to argue that causal mediation methods indeed present a way to go one step further and build a bridge between the causal mediation toolkit and questions in development, and to illustrate how the concept of direct and indirect effects can offer rich insights unavailable via standard reduced-form approaches. We introduce key concepts and estimands in causal mediation, how these relate to the ATE, discuss identification, and present multiple applications. This paper thus makes a systematic attempt to introduce the literature on causal mediation to a development audience, and to outline how these methods can be put to work, which we hope will open up a promising line of research.

The remainder of the paper proceeds as follows. Section 2 provides the motivation behind causal mediation and introduces key concepts: graphical representation of causal structures in the form of Directed Acyclic Graphs (DAGs), and causal estimands of interest such as direct and indirect effects. Section 3 offers a formal discussion of identification and estimation strategies for causal effects and extensions within this framework. Section 4 presents four distinct examples of how causal mediation can be put to work in the context of development, drawing on work from multiple contexts. Section 5 concludes by discussing some challenges and extensions offered by this framework.

2 | MEDIATION: WHY AND WHAT

Starting with the premise that our interest is in uncovering mechanisms behind development outcomes, what is needed is a conceptual framework that (a) enables us to specify these mechanisms with the requisite level of granularity (b) where feasible, allows estimating various causal effects. Field experiments in development economics have largely focused on policy or treatment evaluation, measuring the total effects of a given intervention on the outcome of interest, often through the ATE or similar estimand. This focuses on the question of whether or what works, and the magnitude of the effect. While this is a legitimate question and can be important for programme evaluation, it is not the only question of interest. It is vital to understand why and how these effects ultimately manifest, which calls for investigating the causal pathways through which actual or hypothetical interventions are translated into outcomes in certain populations.

Mediators are intermediate variables that transmit causal effects from actual or hypothetical interventions to outcomes. Causal mediation analysis aims to specify these channels of transmission, and disentangle and estimate the direct and indirect effects that arise from these (Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010; Pearl, 2001; VanderWeele, 2015). A common and effective way to represent causal structures is through the use of Directed Acyclic Graphs or DAGs, which can flexibly accommodate several simultaneous channels through which causal effects manifest.[1] While the traditional approach to causal mediation analysis proposed by Baron and Kenny (1986) specifies linear models for the outcome and mediator and assigns causal interpretation to model coefficients, the modern approach integrates graphical methods with the potential outcomes framework of Rubin (1974) and the pathway analysis pioneered by Sewall Wright (1934).[2] This approach has been largely developed and applied within disciplines including sociology, psychology, epidemiology and political science, but less so in economics and econometrics.[3] The key insights yielded by this framework lie in its ability to define and measure path-specific effects.

Below, we describe different aspects of the causal mediation framework: specifying causal structures, and intermediate variables using DAGs, defining estimands of interest, conditions under which these path-specific effects can be identified, and related estimation strategies. We note that this framework is general enough to accommodate non and semi-parametric approaches.

We can motivate this framework through the following example. Childhood socioeconomic position (SEP) is believed to influence adult cardiovascular mortality. Hossin et al. (2021) posit that this relationship is mediated by social (adult education and SEP) and behavioural factors (smoking, alcohol drinking, physical inactivity, poor diet and body mass index). We discuss this example in greater detail in Section 1, but for now let us consider a simplified version with a single mediator, viz. education. Thus, childhood SEP effects the education level of the individual (as an adult), which in turn effects their cardiovascular (CVD) mortality (within the study period). Childhood SEP also influences cardiovascular mortality through causal pathways other than education. To simplify further, suppose that childhood SEP takes on only two values, viz. high (H) and low (L).

The ATE in this context is the difference in expected counterfactual CVD mortality rates between those with high and low childhood SEP, that is, $\mathbb{E}[\mathrm{CVD}(\mathrm{SEP}=H)]-\mathbb{E}[\mathrm{CVD}(\mathrm{SEP}=L)]$. The corresponding randomised experiment—conceptual in this case—is for childhood SEP to be manipulated. Estimating the ATE using observational data requires adjusting for all possible endogenous variables, for instance location or race since these can affect both childhood SEP as well as adult mortality. In the terminology of mediation, the ATE is known as the Total Causal Effect or TCE, which naturally suggests the existence of other constituent mediated and non-mediated effects. Given Figure 1 and the role of education therein, the indirect effect of childhood SEP via education is the change in mortality due to the change in education that results from SEP switching from H to L. The corresponding direct effect of childhood SEP is the resulting change in mortality while holding education fixed at some given level.

It is worth noting one crucial distinction between the direct effect and the ATE. Both require (conceptually or otherwise) manipulating SEP. But while the direct effect requires holding education fixed so as to isolate the effect of SEP alone, the ATE requires ignoring education as a covariate. Were education to be adjusted for (e.g., by adding it as a control variable in an OLS regression), the resulting ATE estimate would be biased since it will subsume both the true ATE as well as the indirect effect via education, which in this case presumably magnifies the effects of childhood SEP. More generally, it is a well-known principle in the causal inference

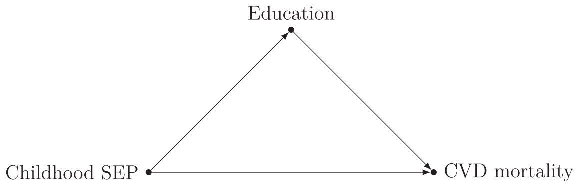

FIGURE 1 Causal structure for effect of childhood socioeconomic position on cardiovascular mortality.

literature that adjusting for intermediate variables (here education) leads to biased estimates of the ATE (Acharya et al., 2016; Gelman et al., 2020). Randomisation—in this case conceptually manipulating childhood SES—can enable estimating the ATE. However, because education is itself effected by childhood SES, randomisation does not in general enable estimating direct or indirect effects via education, a point to which we return below.

3 | PATH SPECIFIC EFFECTS, THEIR IDENTIFICATION AND ESTIMATION

Let us start with a simple causal structure of the type given in Figure 2 with a single treatment variable (A), single mediator (M) and single outcome (Y). This can be extended to accommodate multiple mediators and more complex causal relationships. Several causal effects of interest can be defined. In what follows, M^i_a refers to the value the mediator M would take for a given unit i, when the treatment A is set at $A^i = a$ and likewise $M^i_{a'}$ refers to the value M would take for unit i when $A^i = a'$. Similarly, Y^i_{am} refers to the value Y would take for unit i when $A^i = a$ and $M^i = m$, while $Y^i_{a'M_{a''}}$ refers to the value of Y for unit i when $A^i = a'$ and $M^i = M^i_{a''}$.[4]

The overall treatment effect or TCE for unit i refers to the change in Y^i corresponding to a change in A from $A^i = a$ to $A^i = a'$, encompassing all possible causal pathways—those involving mediators as well as those that operate independently of mediators. Formally, the difference between these two counterfactual outcomes is given by

$$\text{TCE}_i = Y^i_{a'M_{a'}} - Y^i_{aM_a}. \tag{1}$$

Depending on the value of the treatment, only one of $Y^i_{a'M_{a'}}, Y^i_{aM_a}$ can be observed, due to which TCE_i cannot be directly estimated—the fundamental problem of causal inference—and we intend instead to estimate the (average) TCE. Taking expectations, this quantity is given by

$$\text{TCE} = \mathbb{E}\left[Y^i_{a'M_{a'}} - Y^i_{aM_a}\right]. \tag{2}$$

Direct effects are those which sidestep any and all mediators, and reflect changes in the outcome Y due to a change in treatment A while holding mediator M fixed at some level so as to rule out any changes in Y due to a subsequent change in M. The Controlled Direct Effect (CDE)

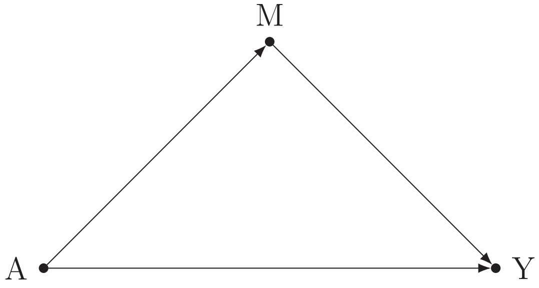

FIGURE 2 Causal structure for effect of treatment A on outcome Y with mediator M.

for unit i refers to the change in Y^i due to a change in A^i while holding M^i fixed—controlled—at some given level m. We can also define natural direct effects, where the word natural conveys that we are measuring the change in Y^i due to a change in A^i when holding M^i at the level it would naturally attain at the reference level of treatment. Depending on the reference levels considered, there are two corresponding natural direct effects. In both cases, we vary the treatment from $A=a$ to $A=a'$. The Pure Natural Direct Effect (PNDE) is defined by holding M^i at M^i_a, while the Total Natural Direct Effect (TNDE) is defined by holding M^i at $M^i_{a'}$. To be clear, in the case of CDE, M^i is held at some $M^i=m$ for all i, whereas for the PNDE and TNDE, holding M^i at the level it would naturally attain at the corresponding levels of treatment implies that those M^i will likely vary across units. Formally

$$\mathrm{CDE}_i = Y^i_{am} - Y^i_{a'm}. \tag{3}$$

$$\mathrm{PNDE}_i = Y^i_{a'M^i_a} - Y^i_{aM^i_a}. \tag{4}$$

$$\mathrm{TNDE}_i = Y^i_{a'M^i_{a'}} - Y^i_{aM^i_{a'}}. \tag{5}$$

Taking expectations, the corresponding (average) CDE, PNDE and TNDE are given by the following expressions

$$\mathrm{CDE} = \mathbb{E}\left[Y^i_{am} - Y^i_{a'm}\right]. \tag{6}$$

$$\mathrm{PNDE} = \mathbb{E}\left[Y^i_{a'M^i_a} - Y^i_{aM^i_a}\right]. \tag{7}$$

$$\mathrm{TNDE} = \mathbb{E}\left[Y^i_{a'M^i_{a'}} - Y^i_{aM^i_{a'}}\right]. \tag{8}$$

For instance, in the example set out above, the CDE could be variously defined depending on the level at which education is conceptually held fixed, while manipulating SEP. For the TNDE (say), we let each individual's education attain the level it naturally would if their SEP = H, and while holding education fixed at this level, estimate the change in average CVD mortality when instead SEP = L.

Indirect effects are those which operate via mediators, or in the simplest case as in Figure 2, a single mediator. These sidestep any changes in Y due to the direct effect of A, thereby focusing exclusively on the ‘knock-on’ effects wherein A effects M and M effects Y. The conceptual manipulation in this case is to vary the mediator from the level it naturally attains when $A=a$ to that attained when $A=a'$ while holding the treatment itself fixed. The level at which the treatment is held gives rise to definitions corresponding to those above. The Pure Natural Indirect Effect (PNIE) for unit i is defined by holding the treatment at $A^i=a$, while the Total Natural Indirect Effect (TNIE) is defined by holding the treatment at $A^i=a'$. In both cases, the causal quantity of interest is the change in Y^i resulting exclusively from the change in $M^i=M^i_a$ to $M^i=M^i_{a'}$.[5,6] Formally

$$\mathrm{PNIE}_i = Y^i_{aM^i_{a'}} - Y^i_{aM^i_a}. \tag{9}$$

$$\mathrm{TNIE}_i = Y^i_{a'M^i_{a'}} - Y^i_{a'M^i_a}. \tag{10}$$

With the corresponding (average) PNIE and TNIE given by the following

$$\mathrm{PNIE} = \mathbb{E}\left[Y^i_{aM^i_{a'}} - Y^i_{aM^i_a}\right]. \tag{11}$$

$$\mathrm{TNIE} = \mathbb{E}\left[Y^i_{a'M^i_{a'}} - Y^i_{a'M^i_a}\right]. \tag{12}$$

For the example above, the PNIE of childhood SEP is the expected change in CVD mortality that arises, holding SEP fixed at SEP $= L$, from the change in education levels from $E_{\mathrm{SEP}=L}$ to $E_{\mathrm{SEP}=H}$. This is arguably a quantity of interest to, say, a policymaker who wants to understand the likely change in mortality due to an intervention that targets education, changing education levels from $E_{\mathrm{SEP}=L}$ to $E_{\mathrm{SEP}=H}$. That is, this quantity compares CVD mortality in two natural states of the world: when everyone attained the education they would with low SEP vs that attained with high SEP, even if the SEP itself stays fixed (in this case at SEP $= L$). Finally, we note that the TCE (Equation 1) of a change from $A = a$ to $A = a'$ can be decomposed as the sum of the PNDE and TNIE, or of the PNIE and TNDE (see Pearl, 2001).

3.1 | Identification

In essence, identifying the various causal effects defined above requires being able to identify the constituent causal relationships, viz. treatment-mediator, mediator-outcome conditional on treatment, and treatment-outcome conditional on mediator. These can be expressed formally as follows, under the additional overall assumption that the data is generated from a Non-Parametric Structural Equation Model (Pearl, 2009, 2014).[7]

A1: Consistency: (i) Consistency of A on M and consistency of $\{A, M\}$ on Y. This assumption requires that actual and potential values coincide for all relevant variables.

A2: Causal effect of treatment on mediator is identifiable: There exists a set of variables W_1 such that the effect of treatment on the mediator can be identified by conditioning on W_1. That is, $M_a \perp\!\!\!\perp A \mid W_1$.

A3: Holding treatment fixed, the causal effect of mediator on outcome is identifiable: There exists a set of variables W_2 such that for each value of the treatment $a, a' \in A$, $Y_{am} \perp\!\!\!\perp \mathrm{M} \mid A = a', \{W_1, W_2\}$. In other words, holding treatment A fixed, the causal effect of the mediator on the outcome can be identified. Note that this rules out (a) unobserved mediator-outcome confounders, and (b) observed mediator-confounders which are themselves affected by the treatment.

A4: Causal effect of treatment on outcome is identifiable: There exists a set of variables W_3 such that $Y_{aM} \perp\!\!\!\perp A \mid \{W_2, W_3\}$. That is, holding M fixed, the effect of treatment on outcome can be identified by conditioning on $\{W_2, W_3\}$.

A significant proportion of the literature uses to a closely related version of these assumptions, usually referred to as sequential conditional independence (see Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010).

Consistency (A1) requires that the functional form relating treatment A to potential outcomes for the mediator M, and that relating $\{A,M\}$ to potential outcomes for outcome Y, are both well-defined.[8] Besides A1, identifying the ATE or TCE alone—as in conventional causal inference—requires only A4, that is, adjusting for all variables that effect both treatment and outcome. Identifying the CDE additionally requires part (a) of A3 in order to identify the effect of the mediator on the outcome. Finally, identifying natural direct and indirect effects also requires identifying the treatment-mediator relationship (A2) and the absence of any confounders of the mediator-outcome relationship that are themselves effected by the treatment, whether or not these are observed, viz. A3 part (b). Randomising the treatment guarantees A4 and A2, while additionally randomising the mediator guarantees A3 part (a). But even if both applications of randomisation are feasible, identifying the natural effects defined above also requires A3 part (b) as well as—crucially—knowledge of the levels that mediators would naturally attain at required levels of treatment. In general therefore, randomisation is not sufficient for identifying natural direct and indirect effects (see Pearl, 2014, p. 460). Alternatively, it is also possible to use quasi-experimental techniques as an alternative to randomisation, for the treatment or mediator or both, such as via instrumental variables or difference-in-difference approaches. While a full discussion is outside the scope of the current paper, this has been comprehensively reviewed in Celli (2022) to which we refer the interested reader for a detailed discussion of the related theoretical and empirical literature.

3.2 | Estimation strategies

Multiple estimation strategies have been proposed in the literature. The traditional, fully parametric approach consists of specifying linear models for the mediator and outcome, and interpreting coefficient estimates in terms of specific causal relationships. As MacKinnon et al. (2020) explain, it is possible to define the direct and indirect effects defined above as functions of coefficients from these models. That is, to obtain equivalent point estimates to specific potential-outcome contrasts, with standard errors obtained via bootstrapping. However, the analytical expressions for the required estimands will need to be re-calculated every time the model specification is changed, and providing analytical expressions for complex or nonlinear models can be challenging.

Instead, the two main approaches used in the literature focus on directly estimating contrasts of average potential outcomes rather than deriving analytical expressions. Under the first approach, the outcome is flexibly modelled as a function of treatment, mediators, relevant interaction terms between these, and the covariates needed to satisfy assumptions A2–A4, and each mediator is likewise modelled as a function of treatment, other mediators where relevant including any interaction terms, and other required covariates. Linear or non-linear models can be specified in this step, with interaction and higher-order terms included as appropriate. Next, corresponding to required levels of the treatment, predicted values for each mediator across the sample are obtained using the model estimates and covariate values for each unit. Finally, the same step is repeated to obtain predicted values of the outcome using the model estimates by holding the treatment at required levels and replacing the mediator(s) with the predicted values obtained in the previous step. The empirical averages of these predictions provide the point estimates for each expectation term in the definitions above.

Under this approach, the sampling variability of the conditional counterfactual distributions of mediators and outcome are accounted for through Monte-Carlo simulation. This can be done in two ways: by taking draws from the sampling distribution of the model predictions (e.g., De Stavola et al., 2015; Vansteelandt & Daniel, 2017), or from the sampling distribution of the model parameters (e.g., Imai, Keele, & Tingley, 2010), and both methods can be combined with bootstrapping. A second approach to estimating direct and indirect effects is through inverse probability weighting as suggested by Huber (2014) who demonstrates this approach in the context of a job training programme. This approach directly estimates the sample analogues to Equations (5)–(10) using two sets of propensity scores for the treatment, conditional on all relevant confounders, and additionally the mediator. Standard errors are again obtained through bootstrapping.

3.3 | Multiple mediators and observed confounders

The definitions provided above can be generalised to more complex causal structures involving multiple mediators, however the associated identification requirements can become significantly more arduous. A generalised version is presented in Figure 3, where not only are there pathways from treatment to mediator to outcome, but also potential causal pathways amongst the mediators themselves. A full treatment of this matter is beyond the scope of this paper, however, a brief outline is as follows.

The simplest case is when all mediators can be treated as a single bloc.[9] Provided assumptions A1–A4 can be met for each mediator individually and also as a set, the TCE, and natural direct and indirect effects can be estimated considering mediators as a group.

Treating mediators other than as a group can pose some challenges. As a simple example, suppose there are only two mediators M_1 and M_2, and that the researcher wants to estimate the natural indirect effect via M_1, that is, a quantity such as $E\left[Y_{aM_{1a}M_{2a}} - Y_{aM_{1a'}M_{2a}}\right]$ where the notation now accounts for two mediators. The first term is straightforward, as it entails holding the treatment at $A = a$ and both mediators at the level they would attain when $A = a$. The second term involves holding the treatment unchanged, letting M_2 attain the value it would at $A = a$, however holding M_1 at the value it would attain when $A = a'$. Identification is still straightforward provided there is no causal arrow from M_2 to M_1. However, if M_2 effects M_1, obtaining the counterfactual $M_{1a'}$ requires setting $A = a'$ and $M_2 = M_{2a'}$, even though $Y_{aM_{1a'}M_{2a}}$ requires

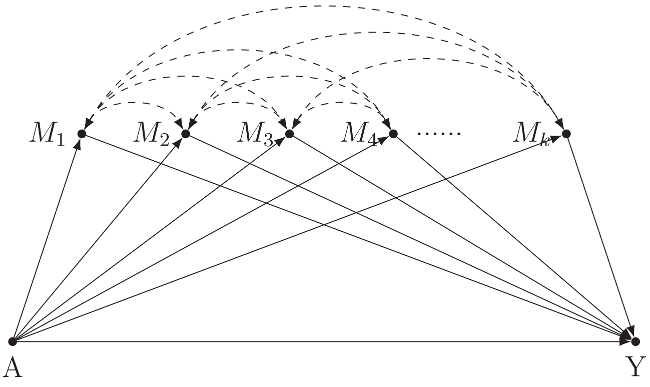

FIGURE 3 Generalised framework with multiple mediators. Solid arrows represent known causal relationships, while dotted bidirectional arrows represent potential causal relationships.

$M_2 = M_{2a}$, where the latter requires setting $A = a$. For this reason, $M_{1a'}$ is termed a 'cross-world' counterfactual, estimating which requires further assumptions about the data generating process.[10]

A final challenge relates to assumption A3 above. This assumption rules out any unobserved mediator-outcome confounders, however in general, we additionally need to rule out any observed mediator-outcome confounders too if they are themselves effected by the treatment: an observed confounder of this sort would be analogous to M_2, viz. a mediator that causally effects a second mediator. De Stavola et al. (2015) discuss a part-solution by way of implementing two alternative, weaker sets of assumptions proposed by Petersen et al. (2006) and Robins and Greenland (1992), while VanderWeele et al. (2014) suggest (a) treating all mediators as a block; (b) focusing on alternative path-specific effects which also yield a decomposition of the TCE; (c) assuming randomisation of the mediator is possible, to estimate a randomised-interventional form of direct and indirect effects.

4 | APPLICATIONS IN DEVELOPMENT

Clearly, the concept of causal mediation is not new, and there exists a large literature on identification and estimation, including several methods by which quantities of interest can be estimated.

However, the extent to which mediation as a conceptual framework is used varies widely across disciplines and areas, and we would argue that the study of development is a field where it has found almost no application. This is in stark contrast to the nature of most developmental processes. Howsoever generally that term is defined, causal explanations in social and economic development require delineating multiple interlinked phenomena. Mediation can be useful in broadening the set of conceptual and empirical tools available to the researcher. We now illustrate this argument with four examples.

4.1 | Female labour force participation

Female labour force participation (FLFP) is a key indicator of economic development. The expectation is that FLFP eventually rises with rising income and education, although it might initially decrease before the eventual increase, giving rise to a U-shaped relationship (Klasen, 2019). However, this hypothesis might play out differently depending on historical, cultural and economic context (Jayachandran, 2021; Klasen et al., 2021).

In its simplest form, FLPF should rise due to the demand for labour and rising productivity due to increases in education. However, at least two factors might counter this. The cultural context might not support the idea of women working outside the household. Second, the nature of marital matching on education might give rise to situations where higher-educated females have higher-educated husbands, resulting in a lower marginal contribution of female earnings to the household. Clearly there are significant factors beyond these as well such as the gender wage gap, demand for higher-educated labour, and regional variation in both.

Conventional approaches assign a causal interpretation to the correlation between education and FLFP, adjusting for factors such as location, regional labour market indicators, household characteristics, and—for cultural contexts where this is relevant—the husband's level of education, earnings, so forth. However, this approach fails to yield causal estimates, because most of the characteristics adjusted for are not confounders; they are in fact mediators. If we conceptualise FLFP as the final outcome and the causal variable of interest to be level of education (E), E effects FLFP via intermediate outcomes: characteristics of the husband following marital matching and the decision of where to live (location and regional labour market conditions). The variables listed above are therefore mediators, which are themselves effected by E, and go on to effect LFP. There might of course also be confounders which need to be added to the model, but these will be in addition to mediators.

Causal mediation affords two benefits here. It yields unbiased estimates of the TCE of education on FLFP, but also enables the researcher to estimate key indirect effects of interest, such as how the nature of marital matching or post-marriage location choices shape FLFP, and how these vary according to categories such as religion or caste.

Kumar and Kao (2022) demonstrate this approach in the context of FLFP in India using data from the second round (2011–2012) of the nationally representative India Human Development Survey (IHDS) (Desai & Vanneman, 2015; Desai et al., 2010). They focus on the simplest case using only a single mediator, viz. the husband's level of education. They find that positive assortative marital matching—women with higher levels of education marry men with higher levels of education, in a context where most marriages are arranged—depresses FLFP because FLFP falls with rising levels of husband's education. The latter might be due to a combination of cultural factors including caste, and the lack of suitable employment opportunities for educated women (Das, 2006; Klasen & Pieters, 2015).

They adjust for relevant covariates in order to identify the three constituent causal relationships discussed above. Given the arranged-marriage context, identifying the causal link between female education and that of the husband requires adjusting for the socio-economic status of the woman's parents, under the assumption that they play the central role in identifying a suitable husband. To this end they adjust for maternal and paternal education, religion and caste, although data on income are unfortunately not available. Next, identifying the causal effect of husband's education on female LFP ideally requires accounting for the social attitudes of the husband and other members of the marital household whose views might shape the wife's LFP—specifically the in-laws. As proxies for these variables, they adjust for in-laws' levels of education, religion and caste. The same covariates are used to identify the third causal link, viz. female education—female LFP while holding fixed the husband's level of education. They follow the algorithm outlined in Section 3.2 to estimate the total and natural direct effect of female education on binary LFP as well as a categorical version of LFP that includes sector, and account for sampling variability using the approach suggested by De Stavola et al. (2015) and Vansteelandt and Daniel (2017). Husbands' education is modelled using OLS, while binary (categorical) LFP is modelled using logit (multinomial logit) models, adjusting for the various covariates discussed. However, they do not model location as a mediator in its own right. Doing so would help explain some of the variation due to differences in the availability of employment opportunities, even if not the absolute levels. Similarly, undertaking this same analysis for subsamples by caste group would yield further insight into the role of cultural factors.

4.2 | Discrimination

Discrimination can accentuate existing inequalities and impede development, and often manifests in the form of differences arising due to group identity such as race, gender, caste or other immutable characteristics. However, unequal outcomes may arise for a variety of reasons and therefore not always be directly attributable to discrimination alone. A key aim of studying discrimination is therefore to distinguish it from other mechanisms by which group-belonging gives rise to inequalities.

Situations involving discrimination can be readily expressed using the language of causal inference and potential outcomes. Often, the question we seek to answer is: ‘Would the outcome(s) be different had the individual been a member of a different group, with everything else remaining the same?’, which suggests that we can define, identify and estimate discrimination in terms of counterfactual contrasts of potential outcomes.

Various forms of discrimination have been identified in the literature. Limited information gives rise to statistical discrimination (Arrow, 1973; Phelps, 1972) when decision-makers cannot observe a certain relevant attribute for the individual (e.g., trustworthiness in a hiring situation) and instead use group-belonging (e.g., race or caste) as a proxy. Instead, prejudice by members of one group against another gives rise to taste discrimination (Becker, 1957). In most real-world situations (e.g., hiring, police stop-and-search, bank lending) we would a priori expect both types of discrimination to operate, but distinguishing between them is key to understanding how discrimination arises and, where relevant, devise remedial measures.[11] Treating discrimination as a problem of causal inference enables examining and estimating both types of discrimination simultaneously, and it is here that mediation offers an enabling framework.

Kumar and Venkatachalam (2021) provide a comprehensive discussion of this problem and a demonstration by way of examining racial discrimination in search-and-frisk actions by the New York Police Department. They define statistical discrimination as the Natural Indirect Effect of group-belonging which operates via beliefs, while taste discrimination is defined the Natural Direct Effect of group-belonging (Pearl, 2001). In their application, statistical discrimination operates via police officers' suspicions about the crime that an individual might have or might be about to commit. These suspicions are the beliefs that drive search-or-frisk decisions. Statistical discrimination arises if these beliefs or suspicions vary by race, and its magnitude equals the corresponding Natural Indirect Effect of race upon outcomes that operates via this particular mediator. Taste discrimination instead reflects racial prejudice alone, bypassing any mediators including beliefs, and therefore equals the Natural Direct Effect of race on search-or-frisk outcomes.

Figure 4 illustrates this causal setup. Race shapes suspicions about criminality, which in turn shape search-or-frisk outcomes, thereby giving rise to statistical discrimination. Race can also shape outcomes directly, reflecting prejudice or taste discrimination. As Kumar and Venkatachalam discuss, there might be multiple mediators of which beliefs are only one, and depending on the causal ordering amongst mediators it can be challenging to estimate statistical

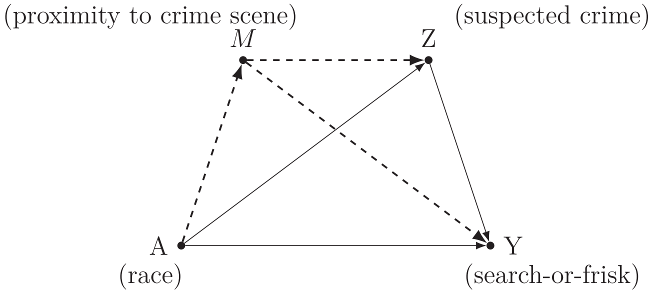

FIGURE 4 Causal structure for discrimination.

discrimination. They demonstrate one such case involving ‘proximity to crime scene’ as a second mediator, which could plausibly influence beliefs about criminality too, thereby presenting estimation challenges of the sort discussed in Section 3.3, in turn necessitating stronger assumptions about causal structure. Their estimation approach combines bootstrapping and draws from the sampling distribution of mediators and outcome following De Stavola et al. (2015) and Vansteelandt and Daniel (2017), and their results show the presence of both types of discrimination against Blacks in police decisions to search-or-frisk. They also suggest an extension to this approach for cases when beliefs are not directly observable, involving Monte-Carlo simulation of a beliefs variable across a range of plausible distributions so as to offer bounds-based estimates of discrimination across a range of scenarios.

Experimental methods can also be used to study discrimination, but have certain limitations. First, the idea of randomising race (or caste, gender) poses a practical challenge. Although it can be partially addressed by instead randomising perceptions of race (e.g., via candidates' names on a CV as in Bertrand and Mullainathan (2004)), however, challenges still remain. In order to estimate taste discrimination, we need to hold fixed both qualifications as well beliefs about trustworthiness. The latter presents a challenge, because beliefs need to be held at the level they would naturally attain, for each individual, corresponding to race being set to Black and White. On the other hand, the ideal experiment for measuring statistical discrimination must isolate the effect of beliefs while ensuring race-based prejudice is held fixed. This requires manipulating information about trustworthiness across belief-levels corresponding to, say, Blacks and Whites, while holding perceived race fixed at Black. The practical unfeasibility of both strategies is a special case of the general problem described by Pearl (2014) wherein randomisation is seldom sufficient to recover natural direct and indirect effects. In contrast, causal mediation offers a unified framework for studying taste and statistical discrimination, which can use observational or a mix of observational and experimental data to estimate discrimination.

4.3 | Political attitudes and influence of institutions

Political science is an outlier in the social sciences in that causal mediation is frequently used to study questions about political preferences and decisions through a mixture of experimental and non-experimental approaches. Several of these questions have clear relevance for topics in development. The examples discussed below span causal processes at individual and macro-levels, illustrating the benefits afforded by causal mediation and clear links to questions in development.

Brader et al. (2008) study the effects of media framing on individuals' attitudes and beliefs. They show how attitudes towards immigration are shaped by exposure to news stories about the costs of immigration, and how this varies according to the race of immigrants featured in these stories. They use an experimental design where racial profiles in the news stories were carefully randomised to ensure no other immigrant-features differed significantly. They find that white individuals react more strongly to news stories featuring Latino immigrants and not for European immigrants, and suggesting that these effects might manifest by triggering certain emotions, show that anxiety is the key mediator whose levels rise in response to these news stories. Imai et al. (2011) revisit this analysis, use graphical methods to exemplify the identification assumptions and various causal estimands of interest. Specifically, since treatment (race in the news story) is randomised, the TCE on immigration attitudes will be identified without further assumptions. However anxiety is a mediator and is not randomised, so that identifying the indirect effect via anxiety is not possible without further assumptions needed in order to identify the anxiety-attitude relationship. Imai et al. (2011) provide a detailed discussion, including on how to use sensitivity analyses to probe the validity of such assumptions and how, when feasible, randomisation of the mediator can be useful in identifying causal effects of interest.

Bormann et al. (2019) draw on Imai et al. (2011) and utilise causal mediation analysis to examine the effects of formal ethnic power-sharing institutions on ethnic conflict. They posit power-sharing behaviour as the key intermediate outcome and thus mediator, and show that institutions do not have a direct effect on the likelihood of peace, but this increases only if actual power-sharing behaviour—the mediator—changes as well. In other words, the TCE can be explained primarily due to the indirect effect. They go on distinguish between two types of conflict: that between members of any power-sharing coalitions, and that between coalition members and non-members. With larger coalitions, the likelihood of infighting rises while that of the second type of conflict decreases. Posing this as a question of causal mediation reflects a crucial attribute and contextual insight: power-sharing behaviour is distinct from power-sharing institutions, and that both shape conflict outcomes together. Their methodology overcomes the shortcomings of employing interaction effects which would not be able to clarify the channels through which institutional arrangements affect conflict. Their estimation strategy is as follows. Using OLS, they model power-sharing behaviour as a function of institutions, and using logit, the onset of ethnic conflict as a function of (power-sharing) institutions and behaviours. This enables disentangling the total effect of institutions structures on conflict into the indirect effect mediated through behaviours and the direct effect consisting of all other mechanisms outside those mediated by power-sharing practices.

4.4 | Policy and programme evaluation

Programme and policy evaluation have been a predominant focus of research efforts in development economics, particularly through randomised control trials. A wide array of questions ranging from the effectiveness of macro policies such as development aid, to micro interventions such as availability of microfinance, hiring extra teachers, and use of fertilisers, so forth have been investigated in the literature using experimental (RCTs, quasi-experiments, natural experiments) as well as observational data. Within this approach, ATEs have been the prime parameter of interest, even as some studies endeavour to go a step further to measure heterogeneous treatment effects.

The focus on ATEs alone is subject to what is now a well-developed critique. As Deaton (2010a, p.424) argues, 'RCT-based evaluation of projects, without guidance from an understanding of underlying mechanisms, is unlikely to lead to scientific progress in the understanding of economic development'. Even within a programme evaluation context, there is need for reliable knowledge not just on what works, but mechanisms that shed light on the why question and identifying the contexts in which projects may be likely to work. Deaton's critique of Urquiola and Verhoogen (2009) offers a case in point on the perils of ignoring mechanisms. Causal mediation methods can help ameliorate some of these issues.

As an example, we consider two evaluations of the effectiveness of caseworkers in enabling unemployed workers to find employment. Caseworkers in employment offices in advanced economies like Switzerland have a dual role to play: they offer counselling and support to those seeking employment, as well as monitoring their job search. They can have a more strict or confrontational style of dealing with the unemployed, or strike a more co-operative and accommodating tone in assigning jobs, sanctions and so on. Using linked caseworker-job-seeker data for Switzerland, Behncke et al. (2010) examine how the level of co-operativeness of caseworkers (elicited via a survey administered to them) effects the employment-probabilities of their clients. They find that less co-operative caseworkers are better at finding employment for their clients. However, focusing solely on ATEs masks the mechanisms that are at play. Huber et al. (2017) go a step further, and decompose the positive ATE associated with less co-operative caseworkers into direct and indirect effects. The indirect effect stems from the assignment to active labour market programmes, while all other causal channels such as the threat of sanctions or the pressure to accept jobs constitute the direct effect. They find that the indirect effect is fact negligible, suggesting that the effectiveness of 'uncooperative' caseworkers stems from aspects of their behaviour in counselling, and not from their effectiveness at assigning clients to suitable labour market programmes. Understanding these mechanisms yields insights concerning how the factors which give rise to this apparent effectiveness might in fact be detrimental in the long run.

5 | DISCUSSION

In this paper, we have argued why causal mediation frameworks are relevant and useful for studying topics in development. We have supported this argument by discussing various applications, and demonstrated how causal questions of several types can benefit from being 'thickened' in this way. This entails moving beyond simple ATE-type queries, and instead positing causal mechanisms of varying complexity and estimating constituent causal effects of interest. However, doing so is not without challenges.

Each of the examples discussed above begins with a causal mechanism specified a priori. This theorised mechanism allows the researcher to draw a DAG, designate treatment, outcome and mediating variables and propose their causal linkages. Doing so requires domain knowledge, drawing on intuition, observation and field experience, as well as prior empirical and theoretical work. The resulting propositions might therefore be naturally subject to disagreement and some ambiguity. Contrast this with a simple average-treatment-effect type of causal query. Here too prior knowledge and insight is required to pose a question that is relevant and useful, but in essence this concerns only two variables and not the mechanism linking them. Instead, the spirit of causal mediation necessitates explicit codification of the causal structure a priori. Once specified, questions of identification and estimation of the parameters of interest are investigated within that structure.

A second challenge is posed by the potential presence of unobserved confounders. Here, unlike for ATEs, randomisation is only a partial solution that cannot fully solve identification challenges if the aim is to estimate natural direct effects. This is a well-known problem in the literature, with several proposals offering sensitivity analysis-based approaches to quantify the magnitude of bias (see Huber, 2020, for a comprehensive overview). In the simpler version of this problem, the confounder is not affected by the treatment. Standard sensitivity approaches involving assumptions on the correlation structure of the confounder and variables of interest (any two of treatment, mediator, outcome) can be employed to obtain bounds on the estimands (e.g., Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010). The more challenging version is when the treatment effects the unobserved confounder. Tackling this requires stronger assumptions specifying the conditional correlations between treatment, confounder and variables of interest (VanderWeele, 2010).

A third, related challenge concerns the number of mediators and the associated complexity this introduces for estimation. This has two components. The first is simply the multiplicity of mediator-wise, path-specific effects. For instance, as Daniel et al. (2015) illustrate, $n = 2$ causally ordered mediators give rise to $4! = 24$ different path specific decompositions of the TCE. A closely related challenge arises from causal interrelationships amongst mediators. Identifying path-specific effects for individual mediators is not always possible, and sensitivity analyses in response to partial identification might need to contend with several degrees of freedom. As the number of sensitivity parameters increases, the combinations of their values requiring investigation increases exponentially, posing challenges not only of computation, but also how to effectively summarise the results.

Nevertheless, we believe that causal mediation offers a rich conceptual framework with several applications in development as well as potential extensions. For instance, intersectionality is the study of differences and inequalities that arise from the intersection of categories such as sex, race and caste and the joint disparities such intersection can give rise to. What adds to the complexity is that frequently, outcomes of interest (e.g., employment) are determined after an intermediate outcome (e.g., education) which itself displays intersectional disparities. Policy-relevant questions can then be posed, such as the extent to which increasing education levels would lessen the intersectional disparities in employment faced by (say) lower-caste females. This fits naturally within a causal mediation framework as a query about the CDE of a two-dimensional treatment (sex, caste) where the mediator (education) is being set to some policy-specified level. A growing literature considers these questions, such as how intersectionality can be thought of as a causal phenomenon, assigning counterfactual interpretations to multiple categories of belonging (Bright et al., 2016; Jackson, 2017; Jackson et al., 2016).

A second extension arises from bringing mediation analysis to panel data settings where at least one of (treatment, mediators, outcome) are longitudinal. There are multiple complexities to deal with. First, how to define path-specific effects in a panel context, because the values at which (say) the treatment is to be held is a single-dimensional vector for the cross-sectional setting, but multidimensional for panels. Additional complexity can arise if for instance the treatment and mediator are both longitudinal, and there are sequential causal relationships of the sort where treatment at $T = 0$ effects the mediator at $T = 0$ which in turn effects treatment at $T = 1$ and so on. Second, how to extend and adapt identification and estimation strategies for panel settings. VanderWeele and Tchetgen Tchetgen (2017) provide a detailed treatment of these questions as part of a nascent but growing literature.

In conclusion, we believe that there is a need to broaden both methods and scope of enquiry within development economics. Causal mediation analysis provides one way of doing so. It preserves and extends the aims of the 'credibility revolution' while addressing some of the leading criticisms related to, what is in our view, the over-emphasis on experimental approaches. It also has the potential to incorporate modern developments in machine learning, thereby improving estimation and computation strategies and allowing for greater complexity. Most importantly, causal mediation helps enrich knowledge of development processes—and not simply outcomes—by bringing the attention back to mechanisms.

ACKNOWLEDGMENTS

The authors are grateful to two anonymous referees for their constructive feedback and comments, and to the journal editor Prof. Reetika Khera and other participants at the workshop on "Money, Methods and Morals: What are Development Economists up to?" for their comments and feedback. The authors also thank Prof. Claudia Bernard for helpful conversations on the topic of intersectionality. The usual disclaimer applies.

CONFLICT OF INTEREST STATEMENT

The authors confirm that there are no relevant financial or non-financial competing interests to report.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Sunil Mitra Kumar https://orcid.org/0000-0002-3959-5238
Ragupathy Venkatachalam https://orcid.org/0000-0002-1190-6321

ENDNOTES

[1] See Chen et al. (2018) and Robins (2003) and see Pearl and Mackenzie (2018) for an accessible introduction.

[2] See Pearl (2001) and VanderWeele (2015) on the use of the potential outcomes framework for causal mediation, and see Celli (2022) on the historic origins of this approach in the context of economics where, albeit, this method is not widely used.

[3] As Imbens (2020, p. 1145) points out, causal mediation has not gained wide acceptance within economics, but it has the potential to help explain causal pathways and perhaps deserves more attention in the discipline.

[4] We provide these definitions for the case of continuous Y and M but they are easily extended to the case when either or both of Y and M are categorical. See for instance VanderWeele (2015).

[5] There is generally no indirect effect analogous to the CDE because this would involve holding M at two essentially arbitrary levels.

[6] Both types of natural direct and indirect effects will be equal (viz. PNDE = TNDE and PNIE = TNIE) if there are no interaction terms AM involving treatment and mediator in the expression that determines the outcome.

[7] See Pearl (2014) and VanderWeele (2015, Ch.2) for details.

[8] This is related to a stronger assumption, the Stable Unit Treatment Value Assumption (Rubin, 1980) or SUTVA that is frequently employed in causal inference. As Rubin (2005) explains, the SUTVA implies two subassumptions: (a) that potential outcomes for each unit should be independent of the treatment assignment to all other units; (b) that there are no multiple or hidden versions of the treatment. It follows from (b) that if there is only one version of the treatment, then actual and potential outcomes coincide, or in other words consistency holds (see also VanderWeele & Hernán, 2013). Thus, consistency is necessary for the SUTVA to hold, but not vice-versa.

[9] See VanderWeele and Vansteelandt (2014) for a detailed explanation of this and related cases.

[10] See Daniel et al. (2015) for an example and implementation.

[11] While these are the two dominant forms of discrimination recognised in the literature, they are not the only forms. For instance, institutions can cause discrimination, where the rules or algorithms by which decisions are made can systematically favour certain groups.

REFERENCES

Acharya, A., Blackwell, M., & Sen, M. (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review*, *110*(3), 512–529.

Arrow, K. (1973). The theory of discrimination. *Discrimination in Labor Markets*, *3*(10), 3–33.

Banerjee, A. V., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Penguin.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Becker, G. (1957). *The economics of discrimination*. University of Chicago Press.

Behncke, S., Frölich, M., & Lechner, M. (2010). Unemployed and their caseworkers: Should they be friends or foes? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(1), 67–92.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*(4), 991–1013.

Bormann, N.-C., Cederman, L.-E., Gates, S., Graham, B. A. T., Hug, S., Strøm, K. W., & Wucherpfennig, J. (2019). Power sharing: Institutions, behavior, and peace. *American Journal of Political Science*, *63*(1), 84–100.

Brader, T., Valentino, N. A., & Suhay, E. (2008). What triggers public opposition to immigration? Anxiety, group cues, and immigration threat. *American Journal of Political Science*, *52*(4), 959–978.

Bright, L. K., Malinsky, D., & Thompson, M. (2016). Causally interpreting intersectionality theory. *Philosophy of Science*, *83*(1), 60–81.

Celli, V. (2022). Causal mediation analysis in economics: Objectives, assumptions, models. *Journal of Economic Surveys*, *36*(1), 214–234.

Chen, B., Pearl, J., & Kline, R. (2018). Graphical tools for linear path models. (Technical Report R-469) Retrieved from UCLA, Department of Computer Science website: https://ftp.cs.ucla.edu/pub/stat_ser/r469.pdf.

Daniel, R. M., Stavola, B. L. D., Cousens, S. N., & Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. *Biometrics*, *71*(1), 1–14.

Das, M. B. (2006). Do traditional axes of exclusion affect labor market outcomes in India?.

De Stavola, B. L., Daniel, R. M., Ploubidis, G. B., & Micali, N. (2015). Mediation analysis with intermediate confounding: Structural equation modeling viewed through the causal inference lens. *American Journal of Epidemiology*, *181*(1), 64–80.

Deaton, A. (2010a). Instruments, randomization, and learning about development. *Journal of Economic Literature*, *48*, 424–455.

Deaton, A. (2010b). Understanding the mechanisms of economic development. *The Journal of Economic Perspectives*, *24*(3), 3–16.

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2–21.

Desai, S., & Vanneman, R. (2015). India Human Development Survey-II (IHDS-II), 2011-12. ICPSR36151-v2. Inter-university Consortium for Political and Social Research [distributor].

Desai, S., Vanneman, R., & National Council of Applied Economic Research, New Delhi. (2010). India Human Development Survey (IHDS), 2005. ICPSR22626-v8. Inter-university Consortium for Political and Social Research [distributor].

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.

Haynes, L., Service, O., Goldacre, B., & Torgenson, D. (2012). *Test, learn, adapt: Developing public policy with randomised controlled trials*. UK Cabinet Office and Behavioural Insights Team.

Hossin, M. Z., Koupil, I., & Falkstedt, D. (2021). Early life socioeconomic position and mortality from cardiovascular diseases: An application of causal mediation analysis in the Stockholm Public Health Cohort. *BMJ Open*, *9*(6), e026258.

Huber, M. (2014). Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Econometrics*, *29*(6), 920–943.

Huber, M. (2020). Mediation analysis. In K. F. Zimmermann (Ed.), *Handbook of labor, human resources and population economics* (pp. 1–38). Springer International Publishing.

Huber, M., Lechner, M., & Mellace, G. (2017). Why do tougher caseworkers increase employment? The role of program assignment as a causal mechanism. *The Review of Economics and Statistics*, *99*(1), 180–183.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334.

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, *105*(4), 765–789.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*, 51–71.

Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *arXiv:1907.07271 [stat]*.

Jackson, J. W. (2017). Explaining intersectionality through description, counterfactual thinking, and mediation analysis. *Social Psychiatry and Psychiatric Epidemiology*, *52*(7), 785–793.

Jackson, J. W., Williams, D. R., & VanderWeele, T. J. (2016). Disparities at the intersection of marginalized groups. *Social Psychiatry and Psychiatric Epidemiology*, *51*(10), 1349–1359.

Jayachandran, S. (2021). Social norms as a barrier to women's employment in developing countries. *IMF Economic Review*, *69*, 1–20.

Klasen, S. (2019). What explains uneven female labor force participation levels and trends in developing countries? *The World Bank Research Observer*, *34*(2), 161–197.

Klasen, S., Le, T. T. N., Pieters, J., & Santos Silva, M. (2021). What drives female labour force participation? Comparable micro-level evidence from eight developing and emerging economies. *The Journal of Development Studies*, *57*(3), 417–442.

Klasen, S., & Pieters, J. (2015). What explains the stagnation of female labor force participation in urban India? *The World Bank Economic Review*, *29*(3), 449–478.

Kumar, S. M., & Kao, Y.-F. (2022). Counterfactual thinking and causal mediation: An application to female labour force participation in India. In R. Venkatachalam (Ed.), *Artificial intelligence, learning and computation in economics and finance*. Springer Nature.

Kumar, S. M., & Venkatachalam, R. (2021). Causal inference for discrimination: An analysis of NYPD stop-and-frisk data.

MacKinnon, D. P., Valente, M. J., & Gonzalez, O. (2020). The correspondence between causal and traditional mediation analysis: The link is the mediator by treatment interaction. *Prevention Science*, *21*(2), 147–157.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence* (pp. 411–420). Morgan Kaufmann Publishers Inc.

Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods*, *19*(4), 459–481.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

Petersen, M. L., Sinisi, S. E., & van der Laan, M. J. (2006). Estimation of direct causal effects. *Epidemiology*, *17*, 276–284.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, *62*(4), 659–661.

Robins, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In P. Green, N. L. Hjort, & S. Richardson (Eds.), *Highly structured stochastic systems* (pp. 70–81). Oxford University Press.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, *3*, 143–155.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, *75*(371), 591–593.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322–331.

Teele, D. L. (Ed.). (2014). *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences*. Yale University Press.

Urquiola, M., & Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, *99*(1), 179–215.

VanderWeele, T., & Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiologic Methods*, *2*(1), 95.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, *21*(4), 540.

VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press.

VanderWeele, T. J., & Hernán, M. A. (2013). Causal inference under multiple versions of treatment. *Journal of Causal Inference*, *1*(1), 1–20.

VanderWeele, T. J., & Tchetgen Tchetgen, E. J. (2017). Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(3), 917–938.

VanderWeele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*, *25*(2), 300–306.

Vansteelandt, S., & Daniel, R. M. (2017). Interventional effects for mediation analysis with multiple mediators. *Epidemiology*, *28*(2), 258–265.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, *5*(3), 161–215.

How to cite this article: Kumar, S. M., & Venkatachalam, R. (2023). Poor economics and its missing mechanisms: The case for causal mediation. *Review of Development Economics*, 1–20. https://doi.org/10.1111/rode.13029