MONOTONIC PATH-SPECIFIC EFFECTS: APPLICATION TO ESTIMATING EDUCATIONAL RETURNS

BY ALEKSEI OPACIC [1,a]

[1] *Department of Sociology, Harvard University,* [a] *aopacic@g.harvard.edu*

Conventional research on educational effects typically either employs a "years of schooling" measure of education, or dichotomizes attainment as a point-in-time treatment. Yet, such a conceptualization of education is misaligned with the sequential process by which individuals make educational transitions. In this paper, I propose a causal mediation framework for the study of educational effects on outcomes such as earnings. The framework considers the effect of a given educational transition as operating indirectly, via progression through subsequent transitions, as well as directly, net of these transitions. I demonstrate that the average treatment effect (ATE) of education can be additively decomposed into mutually exclusive components that capture these direct and indirect effects. The decomposition has several special properties which distinguish it from conventional mediation decompositions of the ATE, properties which facilitate less restrictive identification assumptions as well as identification of all causal paths in the decomposition. An analysis of the returns to high school completion in the NLSY97 cohort suggests that the payoff to a high school degree stems overwhelmingly from its direct labor market returns. Mediation via college attendance, completion and graduate school attendance is small because of individuals' low counterfactual progression rates through these subsequent transitions.

1. Introduction. One of the most resilient social scientific findings across a range of national contexts is the strong association between educational attainment and a variety of life outcomes, including earnings, health, social capital, and family stability (Hout, 2012; Chetty, Deming and Friedman, 2023). Conventionally, researchers have taken one of two approaches to evaluating the social and economic returns to education: the first employs a "years of schooling" measure of educational attainment (Angrist and Krueger, 1991, 1992; Kane and Rouse, 1993; Card, 1994; Ashenfelter and Zimmerman, 1997; Card, 1999, 2001; Angrist and Chen, 2011), while the second dichotomizes attainment as a point-in-time treatment. This latter approach has been particularly influential in the study of the impact of postsecondary attainment on earnings, where the treatment considered is often an indicator for whether an individual has attended, or graduated from, college (Brand and Xie, 2010; Carneiro, Heckman and Vytlacil, 2011; Zimmerman, 2014; Goodman, Hurwitz and Smith, 2017; Smith, Goodman and Hurwitz, 2020; Bleemer, 2022; Mountjoy, 2022).

Despite the important insights this literature has made into establishing the causal effect of educational attainment on important social and economic outcomes, extant work has been inattentive to the sequential process by which people make educational transitions (Mare, 1980).[1] At the end of high school, individuals decide whether or not to enroll in college. Among college enrollees, only 60% receive a BA within six years of initial college entry (Snyder, de Brey and Dillow, 2016), with an even lower proportion for low-income students and students of color (Eller and DiPrete, 2018; Zhou and Pan, 2023). Moreover, amidst higher

Keywords and phrases: causal inference, mediation, sequential ignorability, mediation, education.

[1] I use the term "educational transition" to refer both to vertical transitions (e.g. enrollment at a secondary or tertiary institution), as well as to the attainment of a qualification at a given level (e.g. high school graduation or BA completion).

educational expansion in the US, college graduates must increasingly choose whether to enter the labor market or to enroll in post-graduate education. Increasingly, therefore, educational attainment in the US has become a field of multiple levels with sequential transitions, all of which are independently consequential for individuals' labor market outcomes, and therefore of independent scientific interest.

The sequential nature of educational transitions implies that a causal mediation framework can be employed to study the causal paths by which education's "value-added" occurs. Specifically, we can consider the first transition in a sequence of educational levels of interest as a treatment variable, A, and subsequent transitions as mediators that "transmit" the effects of the treatment and of prior transitions, M_k $(1 \leq k \leq K)$. For example, if we are interested in the total effect of high school completion on earnings, we may ask to what extent this total effect operates indirectly, through the effects of college attendance and college completion (putative mediators) on earnings, or directly, through alternative causal pathways. The insight that the total causal effect of education can be decomposed into its direct and indirect effects opens up a range of important research and policy-oriented questions. For example, tracing to what extent an early-stage educational intervention boosts outcomes such as earnings via its promotion of subsequent educational attainment (its indirect effects), or via earnings directly, would enable policy-makers to discern what drives the intervention's value and to hone subsequent policy (e.g. Hurwitz and Howell, 2014; Sullivan, Castleman and Bettinger, 2019; Castleman, Deutschlander and Lohner, 2020; Bird et al., 2021; Dynarski et al., 2021; Black et al., 2023; Turner and Gurantz, 2024). Relatedly, if the early intervention's effects are heterogeneous across demographic groups, assessing the intervention's direct and indirect effects could guide researchers to aspects of the educational experience that either promote or inhibit upward mobility. Nevertheless, prior empirical approaches are not well-suited to answering these questions: a "years-of-schooling" approach captures the direct effect of each additional year of schooling, while the dichotomous approach conflates the direct and indirect effects.[2]

In this article, I introduce a causal mediation framework for analyzing the effects of educational transitions. For the setting of K (≥ 1) monotonic mediators, I develop a general formula that decomposes the total effect of any level of education into $K + 1$ monotonic path-specific effects (MPSEs): a direct effect net of K subsequent educational transitions, reflecting the path $A \to Y$, and K mutually exclusive "continuation" or gross effects, reflecting the paths $A \to M_1 \to Y$, $A \to M_1 \to M_2 \to Y$, and $A \to M_1 \cdots \to M_K \to Y$. Most importantly, this decomposition exploits a unique characteristic of this empirical setting, in which mediators are characterized by "monotonicity": that is, where an individual's potential $k+1$ mediator value is deterministically zero if that individual's kth mediator value is 0. The resultant decomposition of the ATE into $K+1$ monotonic path-specific effects (MPSEs) can be non-parametrically identified under the assumption of sequential ignorability, which allows for the effect of each educational level to be confounded by a distinct set of (observed) intermediate covariates. I introduce several estimation strategies for my proposed decomposition, including a simple linear model-based regression-with-residuals (RWR) procedure, and a non-parametric estimation strategy based on the efficient influence functions (EIFs) of the target parameters (see Chernozhukov et al., 2017; Kennedy, 2022).

[2]A further strand of literature, especially prominent in labor economics, explores labor market returns to horizontal aspects of differentiation within a given educational level (e.g. college selectivity, as well as specific colleges) or college types (e.g. Cohodes and Goodman, 2014; Goodman, Hurwitz and Smith, 2017; Mountjoy and Hickman, 2021; Chetty, Deming and Friedman, 2023; Eller, 2023). While my proposed framework prioritizes the effects of different levels of education, I discuss in my concluding remarks how the framework could be extended to accommodate multivariate mediators.

This study makes three main contributions. Within the realm of education research, I draw on important work by Heckman, Humphries and Veramendi (2018), who present a similar decomposition of the effect of schooling over the early life course, but differs in two important respects. First, I provide nonparametric definitions, identification results, and estimation strategies for decomposing the total effect of schooling through its direct and indirect components. Second, my decomposition accommodates the presence of a distinct set of observed intermediate confounders for each transition. While one limitation of my approach is that I assume away the presence of *unobserved* confounders for each transition, I propose a sensitivity analysis that assesses the robustness of the results to unobserved confounding, under a set of simplifying assumptions.

More broadly, my framework speaks to the burgeoning field of causal mediation analysis in the social, economic, and health sciences, targeted at assessing the causal pathways by which a treatment affects an outcome. While prior literature overwhelmingly focuses on single-mediator decompositions of the ATE, a growing body of work examines mediation estimands in settings with multiple mediators (Avin, Shpitser and Pearl, 2005; Albert and Nelson, 2011; VanderWeele, Vansteelandt and Robins, 2014; Lin and VanderWeele, 2017; Miles et al., 2017; Steen et al., 2017; Vansteelandt and Daniel, 2017; Miles et al., 2020). In particular, in the case of two causally ordered mediators, Daniel et al. (2015) show that the ATE can be decomposed into multiple path-specific effects (PSEs), and outline the assumptions under which some of these effects are identified. Most recently, Zhou (2022a) generalized this framework to the case of K mediators, establishing a set of identifiable PSEs and introducing several regression-based, weighting, and semiparametric efficient estimators. I extend this literature by examining a special empirical setting where the mediators are monotonic. Compared with traditional mediation-based decompositions, monotonicity faciliates PSE identification under weaker identification assumptions, enables identification of all of the causal paths in question, as opposed to just a strict subset of them, and further permits a finer-grained decomposition. The general decomposition also extends previous literature on mediation under monotonicity which has focused exclusively on the case of a single mediator (e.g. Zhou, 2022b).

Finally, I also contribute to a growing parallel literature that proposes a range of nonparametric, and semi-parametric efficient estimators for alternative mediation estimands, based on the efficient influence functions (EIFs) of the causal quantities of interest (e.g. Miles et al., 2020; Farbmacher et al., 2022; Zhou, 2022a), as well as to closely-related work that proposes semi-parametric efficient estimators for dynamic treatment effects (Lewis and Syrgkanis, 2020; Viviano and Bradic, 2021; Bodory, Huber and Lafférs, 2022).

In the following sections, I first introduce the decomposition for the case of a single intermediate educational transition, before discussing the general case of K intermediate transitions and its identification under the assumption of sequential ignorability (Section 2). In Section 3, I introduce a semiparametric estimation strategy for estimating the proposed decomposition, and in Section 4, I illustrate the proposed framework and methods using data from the National Longitudinal Survey of Youth (NLSY97) cohort. Section 5 concludes.

2. Monotonic Path-Specific Effects.

2.1. *A Single Intermediate Transition.* I first consider the case of a single intermediate educational transition (monotonic mediator). Suppressing subscripts i, let A denote an indicator for high school graduation (the initial educational transition), M_1, an indicator for college attendance (a monotonic mediator or transition), and Y, a binary or continuous outcome of interest such as earnings. A single-transition decomposition thus assesses the educational sequence $A \to M \to Y$: high school graduation$\to$college attendance$\to$earnings. In this way, I treat college attendance as a mediator of the total effect of high school graduation on earnings, in relation to which the total effect of high school graduation can be decomposed into an indirect effect (that "flows through" college attendance), and a direct effect (net of college attendance). Following Heckman, Humphries and Veramendi (2018), I refer to this latter term as the "continuation" value of educational transition A.

Using potential outcomes notation, let $M(a)$ denote an individual's potential value of the mediator if their treatment status were set to a, and let $Y(a,m)$ denote that individual's potential outcome if their treatment and mediator statuses were set to a and m, respectively. I assume, as I formalize in the following section, that sequential transitions are characterized by monotonicity; here, this means that individuals who do not complete high school cannot attend college, or $M(0) = 0$. This sequential nature of educational transitions therefore implies the following set potential outcomes: $\{Y(1), Y(0), Y(1,0), Y(1,1)\}$. Further, since by the composition assumption $Y(a) = Y(a, M(a))$ (VanderWeele and Vansteelandt, 2009), under monotonicity we have that $Y(0) = Y(0, M(0)) = Y(0,0)$, and that

$$Y(1) = Y(1,0) + M(1)[Y(1,1) - Y(1,0)].$$

Thus, the individual total effect of A on Y can be decomposed as

$$Y(1) - Y(0) = Y(1,0) - Y(0,0) + M(1)[Y(1,1) - Y(1,0)]. \tag{1}$$

Since these individual-level quantities are unidentified, I focus on their population-level analogs. Taking the expectation of Equation 1, we obtain the following decomposition of the ATE (see Zhou, 2022b):

$$\begin{aligned}
\text{ATE} &= \mathbb{E}[Y(1) - Y(0)] \\
&= \mathbb{E}[Y(1,0) - Y(0,0)] + \mathbb{E}[M(1)]\mathbb{E}[Y(1,1) - Y(1,0)] + \text{cov}[M(1), Y(1,1) - Y(1,0)] && (2) \\
&= \underbrace{\Delta_0}_{A \to Y} + \underbrace{\pi_1 \Delta_1 + \eta_1}_{A \to M \to Y}. && (3)
\end{aligned}$$

Here, Δ_0 and Δ_1 denote the direct effects of the first and intermediate transitions on the outcome, $A \to Y$ and $M \to Y$, respectively, π_1 denotes the total effect of the first transition on the intermediate transition $A \to M$, and η_1 denotes the covariance between the effect of the initial transition on completion of the second and the effect of the second transition on Y. Specifically, η_1 is positive if those who would attend college given high school completion (i.e., $M(1) = 1$) benefit more from college attendance in terms of their later earnings (i.e., have a larger $Y(1,1) - Y(1,0)$) than those who do not (i.e., $M(1) = 0$), and negative if the opposite is true. Meanwhile, the composite term $(\pi_1 \Delta_1 + \eta_1)$ captures the average indirect effect of the treatment via the intermediate transition ($A \to M \to Y$), comprising the sum of (i) the probability of college enrollment if an individual graduated high school, multiplied by the direct of college enrollment, and (ii) the covariance between college enrollment and its direct effect on earnings.

2.2. *Generalization to K Intermediate Transitions.* I now generalize the approach introduced in the preceding section to the case of K intermediate transitions. As previously, I denote the treatment ("initial transition") of high school graduation by A, and use $M_1, \ldots M_K$

to refer to the K subsequent transitions of interest ("intermediate transitions"), where I assume that all of $M_1, \ldots, M_K$ are binary and that for any $i < j$, M_i temporally precedes M_j. For instance, we may wish to decompose the total effect of high school completion on earnings via college attendance (M_1), college completion (M_2) and graduate school attendance (M_3). Let an overbar denote a vector of variables, such that $\overline{M}_k = (M_1, M_2, \ldots M_k)$ and $\overline{1}_k = (A = 1, M_1 = 1, \ldots, M_{k-1} = 1)$. Further, let $[K]$ denote the set $\{0, 1, \ldots, K\}$. In addition, I denote by X a vector of pretreatment confounders of the effect of $(A, \overline{M}_k)$ on (M_{k+1}, Y), and by $\overline{Z}_k = (Z_1, \ldots Z_k)$ a vector of intermediate confounders that may confound the causal effect of M_k on (M_{k+1}, Y). Using potential outcomes notation, $Y(\overline{1}_k, m_k)$ thus denotes an individual's potential earnings if they completed, possibly contrary to fact, the treatment in addition to $k-1$ intermediate transitions, and then either completed ($m_k = 1$) or did not complete ($m_k = 0$) the kth *intermediate* transition. Similarly, $M_{k+1}(\overline{1}_{k+1})$ denotes an individual's potential value of the $k+1$th intermediate transition were that individual to complete the treatment as well as k prior intermediate transitions. As is standard in the mediation literature, I make the following composition assumption (VanderWeele and Vansteelandt, 2009):

ASSUMPTION 1. Composition: $Y(\overline{1}_k, m_k) = Y(\overline{1}_k, m_k, M_{k+1}(\overline{1}_k, m_k)), \forall k \in [K-1], M_0 \equiv A$.

In words, Assumption 1 states that a person's potential outcome under $(\overline{1}_k, m_k)$ is equal to their potential outcome under $A = 1, \ldots, M_{k-1} = 1, m_k$ and under the value M_{k+1} would naturally take under $A = 1, \ldots, M_{k-1} = 1, m_k$. I also invoke the following constraint on units' potential transition values:

ASSUMPTION 2. Monotonicity: $M_{k+1}(M_k = 0) = 0 \forall k \in [K-1], M_0 \equiv A$.

Informally, Assumption 2 (*monotonicity*) states that an individual's potential $k+1$th transition value is deterministically 0 if that individual fails to complete the prior (kth) transition. It is analogous to a one-sided non-compliance assumption within an instrumental variables (IV) framework, which precludes the presence of both "defiers" as well as "always-takers" principal strata. We can then use this assumption to decompose the ATE of A on Y, which I denote by τ_0. Specifically, let τ_k denote the gross effect of the kth mediator on Y, i.e.,

$$\tau_k = \mathbb{E}[Y(\overline{1}_{k+1}) - Y(\overline{1}_k, 0)],$$

let Δ_0 denote the direct effect of A on Y, and let Δ_k denote the direct effect of the kth mediator on Y, i.e.,

$$\Delta_k = \mathbb{E}[Y(\overline{1}_{k+1}, 0) - Y(\overline{1}_k, 0)].$$

To explicate my approach, note that the gross effect of the kth mediator, τ_k, includes not only the direct effect $M_k \to Y$, net of subsequent educational transitions Δ_k, but also the indirect effects of M_k via subsequent transitions ($M \rightsquigarrow Y$, where a squiggly arrow denotes a combination of multiple paths). This insight motivates us to further decompose τ into its direct and indirect components. Under the composition assumption, τ_k can be decomposed as

$$\tau_k = \Delta_k + \pi_{k+1}\tau_{k+1} + \eta_{k+1}, \tag{4}$$

where

$$\pi_{k+1} = \mathbb{E}[M_{k+1}(\overline{1}_{k+1})],$$
$$\eta_{k+1} = \text{cov}[M_{k+1}(\overline{1}_{k+1}), Y(\overline{1}_{k+2}) - Y(\overline{1}_{k+1}, 0)].$$

For $k = 1, \ldots, K-1$, iteratively substituting equation 4 into the corresponding expression for τ_{k-1} yields

$$\tau_0 = \underbrace{\Delta_0}_{A \to Y} + \sum_{k=1}^{K} \underbrace{(\Pi_{j=1}^{k}\pi_j)\Delta_k + (\Pi_{j=1}^{k-1}\pi_j)\eta_k}_{\theta_k \triangleq A \to M_1 \ldots \to M_k \to Y}, \tag{5}$$

where $\Delta_K = \tau_K$, i.e. Δ_K is a *gross* or continuation effect, since this latter path is a composite one that contains all residual paths omitted in the decomposition (i.e., through educational transitions subsequent to K, if they exist). Thus, the θ_k terms capture how much of the total effect of high school completion flows through each intermediate transition considered (i.e., via college attendance, via college completion, and via graduate school attendance), while Δ_0 captures that portion of the total effect that operates directly, net of the K intermediate transitions considered.

2.3. *Identification.* To identify the causal effects of interest, I rely on a series of sequential ignorability assumptions. While most closely associated with the dynamic treatment effects literature, which rely on observing a complete set of time-varying confounders in order to identify longitudinal effects (see e.g. Lewis and Syrgkanis, 2020; Viviano and Bradic, 2021; Bodory, Huber and Lafférs, 2022), these assumptions can be transferred to a mediation context, given the fact that the mediators of interest are all causally ordered. As will be discussed in the following section, sequential ignorability identification assumptions are distinct from - and in fact weaker than - the assumptions typically employed in studies of causal mediation.

Before proceeding, I introduce the following shorthands. Let $M_0 \triangleq A$ and $M_k = \emptyset \forall k < 0$. In order to estimate the decomposition shown in Equation 5, it suffices to identify the expectation of two types of composite counterfactuals ($Y(\overline{1}_k, m_{k+1})$ and $M_{k+1}(\overline{1}_{k+1})$), as well as covariance terms of the form $\text{cov}[M_{k+1}(\overline{1}_{k+1}), Y(\overline{1}_{k+2}) - Y(\overline{1}_{k+1}, 0)]$ $\forall k \in [K-1]$. I invoke the following three assumptions:

ASSUMPTION 3. Consistency: for any unit, if $A = a$, $Y = Y(a)$; if $(A, \overline{M}_k) = \{\overline{1}_k, m_k\}$, then $Y = Y(\overline{1}_k, m_k)$ $\forall k \in [K]$, and if $(A, \overline{M}_k) = \overline{1}_{k+1}$, then $M_{k+1} = M_{k+1}(\overline{1}_{k+1}) \forall m_{k+1} \in \{0, 1\}, \forall k \in [K-1]$.

ASSUMPTION 4. Sequential ignorability: $(M_1(1), Y(a)) \perp\!\!\!\perp A|X$;
$Y(\overline{1}_k, m_k) \perp\!\!\!\perp \overline{M}_k|X, \overline{Z}_k, \overline{M}_{k-1}$ and $M_{k+1}(\overline{1}_{k+1}) \perp\!\!\!\perp \overline{M}_k|X, \overline{Z}_k, \overline{M}_{k-1}, \forall m_k \in \{0, 1\}, \forall k \in \{1, \ldots, K\}, M_0 \equiv A$.

ASSUMPTION 5. Positivity: $p_{A|X}(a|x) > \epsilon > 0$, $p_{M_k|X,A,\overline{Z}_k,\overline{M}_{k-1}}(m_k|x, a, \overline{z}_k, \overline{m}_{k-1}) > \epsilon > 0$ $\forall k \in [K]$.

Assumption 3 (*consistency*) states that a unit's observed outcome equals its potential outcome under a given treatment sequence. Note that under under the Assumption 1 (Composition), if $Y = Y(\overline{1}_k, m_k)$, then $Y = Y(\overline{1}_k, m_k, M_{k+1}(\overline{1}_k, m_k)) = Y(\overline{1}_k, m_k, M_{k+1}(\overline{1}_k, m_k))$,

$\ldots, M_K(\bar{1}_k, m_k, M_{k+1}(\bar{1}_k, m_k), \ldots M_{K-1}((\bar{1}_k, m_k), M_{k+1}(\bar{1}_k, m_k), \ldots, M_{k-2}(\ldots))$. In plain words, the $K-k$ mediators after mediator k all take their natural levels. Assumption 4 (*sequential ignorability*) is the no unmeasured confounding assumption for the treatment and all mediators. It is considered plausible when sufficient pre-treatment and intermediate covariates $(X, \overline{Z}_K)$ are collected. Finally, Assumption 5 (*positivity*) requires that treatment and mediator assignment is not deterministic. Under Assumptions 3-5, $\mathbb{E}[Y(\bar{1}_k, m_k)]$ and $\mathbb{E}[M_{k+1}(\bar{1}_{k+1})]$ are identified, respectively, as

$$\mathbb{E}[Y(\bar{1}_k, m_k)] = \int_x \int_{\overline{z}_k} \mathbb{E}[Y|x, \overline{z}_k, \bar{1}_k, m_k] \left[\prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \bar{1}_{j-1})\right] dP(x) \tag{6}$$

$$\mathbb{E}[M_{k+1}(\bar{1}_{k+1})] = \int_x \int_{\overline{z}_k} \mathbb{E}[M_{k+1}|x, \overline{z}_K, \bar{1}_{k+1}] \left[\prod_{j=1}^{k} dP(z_j|x, \overline{z}_{j-1}, \bar{1}_{j-1})\right] dP(x) \tag{7}$$

For a proof of the above formulas, see Robins (1986). The covariance (η_k) components in the decomposition are then identified as the "residual" terms such as in Equation 4, which follows directly from the fact that all other components in these equations are identified. Thus, for $k \in \{1, \ldots K\}$, we can identify η_k as

$$\eta_k = \tau_{k-1} - \Delta_{k-1} - \pi_k \tau_k. \tag{8}$$

2.4. *A Comparison with Conventional Mediation Analysis with Multiple Causally Ordered Mediators.* The above decomposition has an analog in the context of a mediation–based decomposition of the ATE with multiple ordered mediators, but differs from conventional mediation analysis in important ways. To illustrate the differences, consider a binary treatment, A, an outcome of interest, Y, and a vector of pretreatment covariates, X, and let $M_1, M_2, \ldots M_K$ denote K causally ordered mediators, assuming that for any $i < j, M_i$ precedes M_j, as above. Moreover, let an overbar denote a vector of variables, so that $\bar{M}_k = (M_1, M_2, \ldots M_k)$,$\bar{m}_k = (m_1, m_2, \ldots m_k)$, and $\bar{a}_k = (a_1, a_2, \ldots a_k)$. Using the potential outcomes notation as above, we can define the following expectation of a nested counterfactual,

$$\psi_{a\bar{a}_k} \triangleq \mathbb{E}\left[Y(a, \bar{M}_k(\bar{a}_k))\right],$$

where $\bar{M}_k(\bar{a}_k) \triangleq \left(\bar{M}_{k-1}(\bar{a}_{k-1}), M_k(a_k, \bar{M}_{k-1}(\bar{a}_{k-1}))\right), \forall k \in [K]$. Under Pearl's (2009) nonparametric structural equation model (NPSEM), Zhou (2022a) demonstrates that the ATE of A on Y can be decomposed into $K+1$ identifiable path-specific effects (PSEs) corresponding to each of the causal paths $A \to Y$ and $A \to M_k \rightsquigarrow Y$ $(k \in [K])$:

$$\text{ATE} = \psi_{\bar{1}} - \psi_{\bar{0}} = \underbrace{\psi_{1,\bar{0}_K,} - \psi_{\bar{0}_{K+1}}}_{A \to Y} + \sum_{k=1}^{K} \underbrace{\left(\psi_{\bar{1}_{k+1},\underline{0}_{k+1}} - \psi_{\bar{1}_k,\underline{0}_{k+1}}\right)}_{A \to M_k \rightsquigarrow Y}. \tag{9}$$

This decomposition holds algebraically when Assumption 2 does not hold (i.e., when the mediators are not monotonic). In contrast, the monotonic characteristic of the proposed decomposition leads to several important differences. First, the PSE decomposition of the ATE in general mediation settings is not algebraically unique, and thus the PSEs defined under alternative decompositions will differ if the effects of the treatment and each mediator vary across levels of the other mediators. In fact, depending on the order in which the paths $A \to Y$ and $A \to M_k \rightsquigarrow Y$ are considered, there are $(K+1)!$ identifiable different ways of decomposing the ATE; the decomposition shown in Equation 9 is just one such decomposition.

Consider the case of two causally dependent mediators. In this setting, the causal pathway $A \to M_2 \rightsquigarrow Y$ can be defined with respect to four different combinations of levels of the treatment and first mediator: under (i) $a=1$ and $M_1(1)$, (ii) $a=1$ and $M_1(0)$, (iii) $a=0$ and $M_1(1)$, or (iv) $a=0$ and $M_1(0)$. By contrast, as a direct consequence of monotonicity, the MPSE decomposition is the unique PSE decomposition of the ATE.

Second, for general PSE decompositions of the ATE, the set of identifiable decompositions is merely a small subset of the total number of decompositions that hold algebraically (see Avin, Shpitser and Pearl, 2005). In particular, the identifiable decomposition does not enable us to disentangle the mediating effects of M_k that are direct (net of subsequent mediators) and indirect (through different combinations of subsequent mediators). For example, in the case of two causally dependent mediators, to assess the mediating role of M_1, only the composite path $A \to M_1 \rightsquigarrow Y = (A \to M_1 \to Y) + (A \to M_1 \to M_2 \to Y)$ is identified. By contrast, mediator monotonicity permits a finer-grained decomposition of the ATE: each PSE is identified. In the case of two causally dependent mediators, for example, the causal path $A \to M_2 \to Y$ is zero, and as a result, each of the paths $A \to Y$, $A \to M_1 \to Y$ and $A \to M_1 \to M_2 \to Y$ are identifiable. Figure 1 illustrates the causal pathways defined and identified under the proposed decomposition in the case of two monotonic mediators.

Finally, the sequential ignorability assumption required to identify the MPSE decomposition is weaker than those required to identify a generic PSE decomposition of the ATE. Specifically, the latter requires Pearl's (2009) non-parametric structural equation model (NPSEM), which stipulates that $\left(M_{k+1}\left(a_{k+1}, \bar{m}_k\right), \ldots M_K\left(a_K, \bar{m}_{K-1}\right), Y\left(a_{K+1}, \bar{m}_K\right)\right) \perp\!\!\!\perp M_k\left(a_k, \bar{m}^*_{k-1}\right) \mid X, A, \bar{M}_{k-1}, \forall k \in [K]$. This assumption, sometimes referred to as the "cross-world" independence assumption, is stronger than the sequential ignorability assumption (4) required to identify the MPSE decomposition since it rules out the existence of confounders of the mediators, be they observed or unobserved. By contrast, the MPSE decomposition identification results accommodate *observed* intermediate confounding without altering the substance of the decomposition.

3. Semiparametric, EIF-Based Estimation. The identification results outlined above suggest that the proposed decomposition can be estimated via several approaches, including outcome-based modeling, models for the treatment and mediators via inverse probability weighting, as well as doubly robust approaches. Parametric procedures are attractive because of their conceptual simplicity and ease of implementation: in Supplementary Material A I show how, under a set of linear models for the outcome and mediators, the θ_k components in Equation 5 can be read off from simple functions of coefficients in these linear models. However when X and $\overline{Z}_K$ are high-dimensional, parametric estimators which require a user-defined specification of the data-generating process may suffer biases resulting from model misspecification. In order to reduce model dependency, in this section I provide a nonparametric estimation approach which draws on a debiased machine learning (DML) approach. My DML approach is characterized by two components: first, the use of a Neyman orthogonal estimating equation based on the efficient influence function (EIF) for the target parameters, which makes estimates of the parameter "locally robust" to estimates of the nuisance functions; second, the use of a K-fold cross-fitting algorithm (Chernozhukov et al., 2017).

Let $O = (X, A, \overline{Z}_K, \overline{M}_K, Y)$ denote the observed data, and $\mathcal{P}$ a nonparametric model over O wherein all laws satisfy the positivity assumption described in Section 2. Before proceeding, I define the following auxiliary functions, as introduced in Section 2: $\psi_{km_k} \triangleq \mathbb{E}[Y(\overline{1}_k, m_k)]$ and $\phi_k \triangleq \mathbb{E}[M_{k+1}(\overline{1}_{k+1})]$, for all $k \in [K], M_0 \triangleq A$. Using the identification results given in Section 2, ψ_{km_k} can be written in terms of expectations of observed data:

$$\psi_{km_k} = \mathbb{E}_X \mathbb{E}_{Z_1|X,1} \ldots \mathbb{E}_{Z_k|X,\overline{Z}_{k-1},\overline{1}_k} \mathbb{E}[Y|X, \overline{Z}_k, \overline{1}_k, m_k]. \tag{10}$$

For each $j \in [k]$, we can thus define $\mu^k_{jm_k}\left(X, \bar{Z}_k\right)$ iteratively as

$$\mu^k_{km_k}\left(X, \bar{Z}_k\right) \triangleq \mathbb{E}\left[Y \mid X, \bar{Z}_k, \bar{1}_k, m_k\right],$$
$$\mu^k_{jm_k}\left(X, \bar{Z}_j\right) \triangleq \mathbb{E}\left[\mu^k_{j+1m_k}\left(X, \bar{Z}_{j+1}\right) \mid X, \bar{Z}_j, \bar{1}_{j+1}\right] \forall j \in [k-1].$$

Further, let $\pi_{km_k}\left(X, \overline{Z}_k\right) \triangleq \Pr[M_k = m_k \mid X, \overline{Z}_k, \bar{1}_k] \forall k \in [K]$, and $\pi_{01}\left(X\right) \triangleq \Pr[A = 1|X]$. The efficient influence function (EIF) of ψ_{km_k} is closely related to the EIF for the g-formula, and can be written as

$$\psi_{km_k}(O) = \sum_{j=0}^{k+1} \varphi_j(O), \tag{11}$$

where

$$\varphi_0(O) = \mu^k_{0m_k}(X) - \psi_{km_k}$$
$$\varphi_j(O) = \frac{A}{\pi_{01}\left(X\right)} \left(\prod_{l=1}^{j-1} \frac{M_l}{\pi_{l1}\left(X, \overline{Z}_l\right)}\right) \left(\mu^k_{jm_k}\left(X, \overline{Z}_j\right) - \mu^k_{j-1m_k}\left(X, \bar{Z}_{j-1}\right)\right), \quad j \in \{1 \ldots, k\}$$
$$\varphi_{k+1}(O) = \frac{A}{\pi_{01}\left(X\right)} \left(\frac{\mathbb{I}(M_k = m_k)}{\pi_{km_k}\left(X, \overline{Z}_k\right)} \prod_{k=1}^{k-1} \frac{M_l}{\pi_{l1}\left(X, \overline{Z}_l\right)}\right) \left(Y - \mu^k_{km_k}\left(X, \bar{Z}_k\right)\right).$$

For a proof, see Rotnitzky, Robins and Babino (2017). The semiparametric efficiency bound for any asymptotically linear estimator of ψ_{km_k} in $\mathcal{P}$ is therefore $\mathbb{E}[\left(\varphi_{km_k}(O)\right)^2]$. The EIF motivates an EIF-based estimator for ψ_{km_k}, obtained by solving the empirical moment condition $\mathbb{P}_n[\varphi_{km_k}(O; \hat{\eta})] = 0$, where $\mathbb{P}_n[\cdot]$ denotes an empirical average, and where $\varphi_{km_k}(O; \hat{\eta})$ denotes the estimated EIF, evaluated using plug-in estimators for the nuisance functions. Specifically,

$$\begin{aligned} \hat{\psi}^{\text{eif}}_{km_k} = \mathbb{P}_n\Bigg[& \frac{A}{\hat{\pi}_{01}\left(X\right)} \left(\frac{\mathbb{I}(M_k = m_k)}{\hat{\pi}_{km_k}\left(X, \overline{Z}_k\right)} \prod_{k=1}^{k-1} \frac{M_l}{\hat{\pi}_{l1}\left(X, \overline{Z}_l\right)}\right) \left(Y - \hat{\mu}^k_{km_k}\left(X, \bar{Z}_k\right)\right) \\ & + \sum_{j=1}^{k} \frac{A}{\hat{\pi}_{01}\left(X\right)} \left(\prod_{l=1}^{j-1} \frac{M_l}{\hat{\pi}_{l1}\left(X, \overline{Z}_l\right)}\right) \left(\hat{\mu}^k_{jm_k}\left(X, \overline{Z}_j\right) - \hat{\mu}^k_{j-1m_k}\left(X, \bar{Z}_{j-1}\right)\right) \\ & + \hat{\mu}^k_{0m_k}(X)\Bigg]. \end{aligned} \tag{12}$$

A similar EIF-based estimator can be used for ϕ_k to estimate the π_k terms in Equation 5. This estimator is based on the following nuisance functions for estimation (see Supplementary Material I for details):

$$\gamma_k\left(X, \bar{Z}_k\right) \triangleq \mathbb{E}\left[M_{k+1} \mid X, \bar{Z}_k, \bar{1}_{k+1}\right],$$
$$\gamma_j\left(X, \bar{Z}_j\right) \triangleq \mathbb{E}\left[\gamma_{j+1}\left(X, \bar{Z}_{j+1}\right) \mid X, \bar{Z}_k, \bar{1}_{j+1}\right] \forall j \in [k-1].$$

Next, following Kennedy (2022, p. 15), let $\mathbb{IF} : \Psi \to L_2(\mathbb{P})$ denote the operator mapping the functionals $\{\Delta_k, \pi_k, \eta_k\} : \mathcal{P} \to \mathbb{R},\ \forall \in [K]$ to their respective influence functions under the nonparametric model $\mathcal{P}$. Because the (Δ_k, τ_k) components of the decomposition are linear in ψ_{km_k}, by linearity of the EIF, $(\mathbb{IF}(\Delta_k), \mathbb{IF}(\tau_k))$ can be expressed as linear combinations of $\varphi_{km_k}(O)$. In particular, $\mathbb{IF}(\tau_k) = \varphi_{(k+1)1}(O) - \varphi_{k,0}(O)$ and $\mathbb{IF}(\Delta_k) = \varphi_{(k+1)0}(O) - \varphi_{k,0}(O)$. The EIFs of η_k and θ_k, $\forall k \in [K]$ under $\mathcal{P}$ are derived as in Theorem 3.1:

THEOREM 3.1. *The EIFs of η_k, θ_k $\forall k \in [1, \ldots, K]$ under P are given, respectively, by*

$$\mathbb{IF}(\eta_k) = \mathbb{IF}(\tau_{k-1}) - \mathbb{IF}(\Delta_{k-1}) - \tau_k \mathbb{IF}(\pi_k) - \pi_k \mathbb{IF}(\tau_k),$$

$$\mathbb{IF}(\theta_k) = \mathbb{IF}(\Delta_k) \prod_{j=1}^{k} \pi_j + \Delta_k \sum_{j=1}^{k} \mathbb{IF}(\pi_j) \prod_{\substack{l=1 \\ l \neq j}}^{k} \pi_l + \mathbb{IF}(\eta_k) \prod_{j=1}^{k-1} \pi_j + \eta_k \sum_{j=1}^{k-1} \mathbb{IF}(\pi_j) \prod_{\substack{l=1 \\ l \neq j}}^{k-1} \pi_l,$$

for $k \in \{1, \ldots K\}$, with $\theta_0 = \Delta_0$, and where $\mathbb{RIF}(\phi) = \mathbb{IF}(\phi) + \phi$, denotes the recentered EIF of a parameter (about the truth). Their corresponding EIF-based estimators are (see Supplementary Material I for derivations):

$$\hat{\eta}_k^{eif} = \widehat{\mathbb{RIF}}(\tau_{k-1}) - \widehat{\mathbb{RIF}}(\Delta_{k-1}) - \hat{\tau}_k \widehat{\mathbb{RIF}}(\pi_k) - \hat{\pi}_k \widehat{\mathbb{RIF}}(\tau_k) + \hat{\pi}_k \hat{\tau}_k,$$

$$\hat{\theta}_k^{eif} = \widehat{\mathbb{RIF}}(\Delta_k) \prod_{j=1}^{k} \hat{\pi}_j + \hat{\Delta}_k \sum_{j=1}^{k} \widehat{\mathbb{RIF}}(\pi_j) \prod_{\substack{l=1 \\ l \neq j}}^{k} \hat{\pi}_l + \widehat{\mathbb{RIF}}(\eta_k) \prod_{j=1}^{k-1} \hat{\pi}_j + \hat{\eta}_k \widehat{\mathbb{RIF}}(\pi_j) \prod_{\substack{l=1 \\ l \neq j}}^{k-1} \hat{\pi}_l$$

$$- k \hat{\Delta}_k \prod_{j=1}^{k} \hat{\pi}_j - (k-1) \hat{\eta}_k \prod_{j=1}^{k-1} \hat{\pi}_j.$$

where $\widehat{\mathbb{RIF}}(\phi) = \widehat{\mathbb{IF}}(\phi) + \phi$, and $\widehat{\mathbb{IF}}(\phi)$ denotes the influence function of a parameter evaluated at estimates of its component nuisance functions (see Supplementary Material I for derivations).

When machine learning estimators are used to compute the nuisance functions, in order to ensure the convergence rates outlined in Theorem 3.2 below, one could assume Donsker-type conditions for the nuisance function estimators, which restricts the set of estimators available to use. Alternatively, to expand the class of estimators that can be used for estimating the nuisance functions, sample-splitting can be used. In particular, Chernozhukov et al. (2017) suggest a "cross-fitting" procedure, which comprises the following steps: (1) Randomly split data into J folds: $\{S_1, ...S_J\}$; (2) For each fold S_j, use the remaining $(j-1)$ folds (training sample) to fit a flexible machine-learning model for each of the nuisance functions involved in the estimating equations; (3) For each observation in j (estimation sample), use estimates of the above models to construct a set of estimated RIF functions for $\Delta_k \forall k \in \{0, \ldots, K-1\}$, and for $(\pi_k, \tau_k, \eta_k, \theta_k) \forall k \in [K]$; (4) Compute an estimate of the decomposition components by averaging the estimated RIF functions across all subsamples S_1 through S_J. When all nuisance functions are estimated via data-adaptive methods and cross-fitting, the semiparametric efficiency of θ_k^{rEIF} is given in the following Theorem:

THEOREM 3.2 (Semiparametric efficiency). *Under Assumption 5, and under suitable regularity conditions (e.g. Chernozhukov et al., 2018), then $\hat{\theta}_k^{eif}$ is semiparametric efficient* *if* $\sum_{j=k}^{k+1}\left[\sum_{l=0}^{j} R_n(\hat{\pi}_{l1})R_n(\hat{\mu}_{l0}^{j})\right] + \sum_{j=0}^{k-1}\left[R_n(\hat{\pi}_{j1})R_n(\hat{\mu}_{j0}^{k-1}) + R_n(\hat{\pi}_{j1})R_n(\hat{\mu}_{j1}^{k-1})\right] + \sum_{j=0}^{k}\left[\sum_{l=0}^{j} R_n(\hat{\pi}_{l1})R_n(\hat{\gamma}_{l}^{j})\right] = o(n^{-1/2})$, *where* $R_n(\,\cdot\,)$ *denotes a mapping from a nuisance function to its* $L_2(P)$ *convergence rate, and where* $\hat{\mu}_{l0}^{K+1} \triangleq \hat{\mu}_{l1}^{K}$.

To gain some intuition for the result in Proposition 3.2, we can focus on $\theta_1 = \pi_1\Delta_1 + \eta_1$, i.e., the MPSE through M_1 when $K = 1$. Note that estimation of $\theta_1 = \pi_1\Delta_1 + \eta_1$ requires estimating the following decomposition components: $(\pi_1, \Delta_1, \tau_{0,}, \Delta_0, \tau_1)$. To estimate these components, it suffices to estimate the following quantities: $(\phi_1, \psi_{01}, \psi_{00}, \psi_{10}, \psi_{11})$. In order for $\hat{\theta}_1^{\text{eif}}$ to be semiparametric efficient, we require that the estimators employed for the set $(\phi_1, \psi_{01}, \psi_{00}, \psi_{10}, \psi_{11})$, i.e., $(\hat{\phi}_1^{\text{eif}}, \hat{\psi}_{01}^{\text{eif}}, \hat{\psi}_{00}^{\text{eif}}, \hat{\psi}_{10}^{\text{eif}}, \hat{\psi}_{11}^{\text{eif}})$, are themselves semiparametric efficient. Thus, a sufficient (but not necessary) condition in order for $\hat{\theta}_1^{\text{eif}}$ to obtain the semiparametric efficiency bound is if, for any two nuisance functions involved in $(\hat{\phi}_1^{\text{eif}}, \hat{\psi}_{01}^{\text{eif}}, \hat{\psi}_{00}^{\text{eif}}, \hat{\psi}_{10}^{\text{eif}}, \hat{\psi}_{11}^{\text{eif}})$, the product of their convergence rates is $o(n^{-1/2})$. In this way, $\hat{\theta}_1^{\text{eif}}$ will obtain the semiparametric efficiency bound if all of its constituent nuisance functions converge at a rate faster than $n^{-1/4}$ (although it will also obtain the efficiency bound under a variety of alternative conditions).

When data-adaptive methods are used to estimate the nuisance functions, inference on all components can be conducted via the variance of the empirical analog of the EIF, i.e. $\mathbb{P}_n[(\hat{\psi}_{km_k}^{\text{EIF}})^2]/n$. For example, inference on τ_1 can be conducted by estimating $\mathbb{P}_n[(\hat{\psi}_{11}^{\text{EIF}} - \hat{\psi}_{10}^{\text{EIF}})^2]/n$.

4. Empirical Analysis. To illustrate my approach empirically, I draw on data from the National Longitudinal Survey of Youth 1997 (NLSY97). I parse out the direct effect of high school graduation on adult earnings and its indirect or continuation effects via (i) college attendance, (ii) college graduation, and (iii) graduate school attendance. My analytic sample comprises $N = 7,305$ respondents.

I construct four types of variables: educational transitions, adult earnings, a set of confounders for the effect of high school graduation on subsequent transitions and earnings, and a single set of intermediate confounders for the effect of college completion on subsequent transitions and earnings. My educational transition variables contain a binary treatment denoting whether a respondent had graduated high school by age 22, and three binary mediators denoting whether the respondent had attended a 4-year college by age 22, whether the respondent had received a BA degree by age 29, and whether the respondent had enrolled in a graduate level program by age 29, respectively. I assume that all individuals who make a given educational transition have made all previous educational transitions. Thus, by construction, my coding strategy disallows for cases which violate the monotonicity assumption.[3] My outcome of interest is logged average annual earnings at ages 32-36, which I define to be the (logged) average of a respondent's self-reported wage, salary income, and business income. Earnings are adjusted for inflation to 2023 dollars using the personal consumption expenditures (PCE) index. After dropping respondents with missing earnings information,

[3]Assuming away cases in which an individual makes a particular educational transition without having made *all* previous transitions serves as a reasonable approximation to reality. Among the set of individuals who have non-missing earnings information in the NLSY97 (i.e., those who comprise my analytic sample), 94% of individuals observed to attend graduate school by age 29 also completed a BA by age 29; 93% of respondents who completed a BA by age 29 had attended a 4-year college by age 22 (6% of those who completed a BA by age 29 first attended a 4-year college between ages 23 and 26 inclusive), and 99% of respondents who attended a 4-year college by age 22 had also completed high school.

I accommodate those with zero earnings by adding a small constant of $\$1,000$ to observed earnings (though in Supplementary Material F, I replicate my main analyses under alternative definitions of earnings).

In an effort to satisfy the sequential ignorability assumption (Assumption 4), I include a large array of covariates in my models. This set of covariates is more expansive than those used in previous, observational studies of returns to education (see in particular Scott-Clayton and Wen, 2019). In particular, in addition to including information on respondent demographics (gender, race, ethnicity, age in 1997), and observed pre-college performance such as overall high school GPA and test score on the Armed Services Vocational Aptitude Battery (ASVAB), I include detailed information on socioeconomic background. Since my proposed decomposition also facilitates the inclusion of a distinct set of observed intermediate confounders for each transition, I include two postsecondary characteristics (Z) to adjust for confounders of the effect of BA completion and graduate school attendance on earnings: field of study and college GPA. To assess the robustness of my main conclusion to forms of unobserved confounding, in Supplementary Material C, I produce a set of "bias-corrected" estimates of the decomposition components under certain assumptions about the nature of the confounding.

A large proportion (just under 50%) of respondents are missing information on covariates X and Z. For my main analyses, I impute missing values on these covariates via multiple imputation to increase efficiency, but in Supplementary Material E, I replicate these analyses restricted to the sample of respondents with complete information. This exercise produces substantively similar results (for covariate means for each of these analytic samples, see Supplementary Material D). After constructing the analytical sample, I apply both the DML estimator described in Section 3 as well as a parametric, regression-with-residuals (RWR) algorithm (described in Supplementary Material A) to implement the proposed decomposition. For the DML approach, I estimate all nuisance functions, using a super learner composed of the Lasso and random forest and, following Chernozhukov et al. (2017), use five-fold cross-fitting. All weights involved in computing the rEIFs are censored at their $1st$ and $99th$ percentiles. Supplementary Material H gives further details about the particular models required given my assumed data generation process.

Figure 2 shows my estimates of the average total effect (ATE) on log earnings and its direct and continuation components under both the DML and RWR procedures. Both procedures return similar estimates, though deviate in the estimated magnitude of MPSE θ_1, and DML estimates come expectedly with a significantly greater amount of precision. The first column shows that the estimated ATE of graduating high school on log earnings under DML (RWR) is 0.67 (0.63), which implies an earnings premium of approximately 96%. The next two columns indicate that the vast majority (69% under DML and 75% under RWR) of the ATE operates directly, i.e. net of college attendance, BA completion and graduate school attendance (MPSE θ_0, $A \to Y$). Specifically, high school graduates who do not proceed to college can be expected to earn on average 0.46 (0.47) log earnings more than high school non-completers under DML (RWR), an earnings premium of 59%.

While the majority of the ATE is explained by the direct effect, a non-trivial portion occurs through mediation effects through later transitions. Under DML, the continuation effects of high school graduation via college attendance without BA completion (MPSE θ_1, $A \to M_1 \to Y$) and via BA completion without graduate school participation (MPSE θ_2, $A \to M_1 \to M_2 \to Y$) both mediate roughly 15% of the ATE, and correspond to an earnings premium of approximately 10%. The RWR estimate of θ_1 is notably lower at 0.03 and is also imprecisely estimated. Under both estimation procedures, the continuation effect via graduate school attendance ($A \to M_1 \to M_2 \to M_3 \to Y$) is very small and fails to reach conventional levels of significance. In sum, the total effect of high school graduation on earnings is determined overwhelmingly by its direct effect on earnings.

Table 1 shows DML and RWR estimates of the various components (the direct effects (Δ_k), probabilities (π_k) and covariance terms (η_k)) that constitute the continuation effects θ_k. Several points are of note. First, the components in the table offer insights into the economic and educational returns to different educational stages. The direct effects of each educational transition (Δ_k) are highly variable: they are largest for high school graduation and for college completion (both at 0.46 under DML), and lowest for college attendance and graduate school participation (at 0.2 and 0.12, respectively, under DML). Note that the payoff to graduate school attendance could be depressed by the fact that I observe individuals at a maximum age of only 36, if graduate school earnings premia materialize only much later in the life course. The counterfactual continuation probabilities (π_k) also provide insight into barriers in educational participation. In particular, even if an individual were to complete high school (possibly contrary to fact), that individual would have under a 50% chance of continuing to a 4-year college without further intervention to increase individuals' college application, admissions and enrollment rates. Further, even if individuals were to counterfactually both complete high school and attend a 4-year college, only a very small proportion ($\pi_1 \cdot \pi_2 = 0.24$) would be expected to complete their BA degree without further intervention at the college-level.

Second, the fine-grained nature of the MPSE decomposition enables us to trace the continuation effects to their constituent components. In particular, while the direct effect of high school completion is comparable to the direct effect of BA graduation on earnings, suggesting an earnings premium of 59% relative to college attendance without completion, the continuation effect via BA completion that it informs (MPSE θ_2, $A \to M_1 \to M_2 \to Y$) only mediates a small amount of the overall ATE because because θ_2 is approximately (plus the small value of η_2) equal to Δ_2 scaled by the product $\pi_1 \cdot \pi_2 = 0.24$. In words, despite the relatively large direct effect of BA completion on earnings, given individuals' low counterfactual probability of BA completion, this transition is not an important mediating pathway of the total effect of high school completion on earnings. The result is that college attendance without completion mediates high school graduation's earnings effects as much as BA completion, despite the fact that college attendance without completion yields a much smaller earnings return for high school graduates than BA completion without graduate school attendance does for college enrollees.

One instructive point of comparison for these results are instrumental variable (IV) estimates of returns to years of schooling, typically estimated in the range of 6% to 12% (Angrist and Krueger, 1991, 1992; Kane and Rouse, 1993; Card, 1994; Ashenfelter and Zimmerman, 1997; Angrist and Chen, 2011). While my estimate of the overall return to high school graduation (τ_0) could appear large in this light, several factors could reconcile this difference. First, τ_0 captures the direct *and* continuation effects of high school completion (whereas IV estimates of schooling returns capture schooling's direct effects). Further, τ_0 captures the effect of multiple additional years of schooling (as the high school graduates and high school non-completers that form the comparison group differ by multiple years of schooling), as opposed to a single year's additional return. In fact, we can more directly compare my DML estimate of the direct return to high school graduation (Δ_0) of 0.46 (corresponding to an earnings premium of 58%) using the fact that, in the NLSY97, high school non-completers attained on average 3.7 fewer years of schooling than high school completers. An IV estimate of 12%, for example, would therefore imply an earnings return to 3.7 additional years of approximately 52% - broadly in line with my result. Still, to assess the robustness of the above findings to potential violations of Assumption 4 (Sequential Ignorability), I implement a sensitivity analysis in Supplementary Material C. Under the stated assumptions about the pattern of unobserved confounding, my primary finding that the ATE of high school graduation is overwhelmingly mediated via its direct effect remains highly robust to unobserved confounding.

5. Conclusion. In this article, I have developed a causal mediation framework for analyzing education effects on earnings. First, I have demonstrated that the total effect of any level of education can be decomposed into a direct effect and K mutually exclusive "continuation" effects. All of these effects are identifiable under the assumption of sequential ignorability. Importantly, this property allows for the effect of each educational transition to be confounded by a distinct set of observed covariates - a property which allows for weaker identification conditions compared with conventional mediation-based decompositions of the ATE (Miles et al., 2017; Zhou, 2022a).

Although my empirical motivation is the estimation of educational returns, the proposed framework applies widely to a range of demographic and organizational settings characterized by "state dependency" between treatment and mediators. This characteristic is particularly salient in demographic phenomena, which often involve sequential transitions over the life course. Certain demographic events are rigid in their monotonicity as a result of their definition. For example, researchers may be interested in discerning the degree to which positive effects of marriage on outcomes such as earnings and life satisfaction are undermined by the negative effects of divorce and separation (and, in turn, their mitigation via re-marriage) (Kenney, 2004; Sweeney and Phillips, 2004). Divorce can be "attained" only by individuals who are already married. Similarly, the effect of parenthood on earnings can be seen as operating directly, through the effect of having a first child net of subsequent children, as well as operating indirectly through the effects of having multiple children, transitions which are clearly monotonic in nature. A similar perspective may be taken in a criminal justice context: the total effect of early-stage police contact (such as being searched for contraband) on educational and socio-psychological outcomes can be decomposed into path-specific effects via subsequent arrest and incarceration (Weaver and Lerman, 2010; Kirk and Sampson, 2013; Sugie and Turney, 2017).

Finally, although in this paper I have considered a decomposition of the average treatment effect for the case of binary monotonic mediators, as shown in Supplementary Material G, the framework could straightforwardly be extended to accommodate categorical transitions. Given the heterogeneity of higher-educational trajectories in the US, such an extension would prove useful for modeling the relative payoffs to distinct educational pathways.

6. Tables and figures.

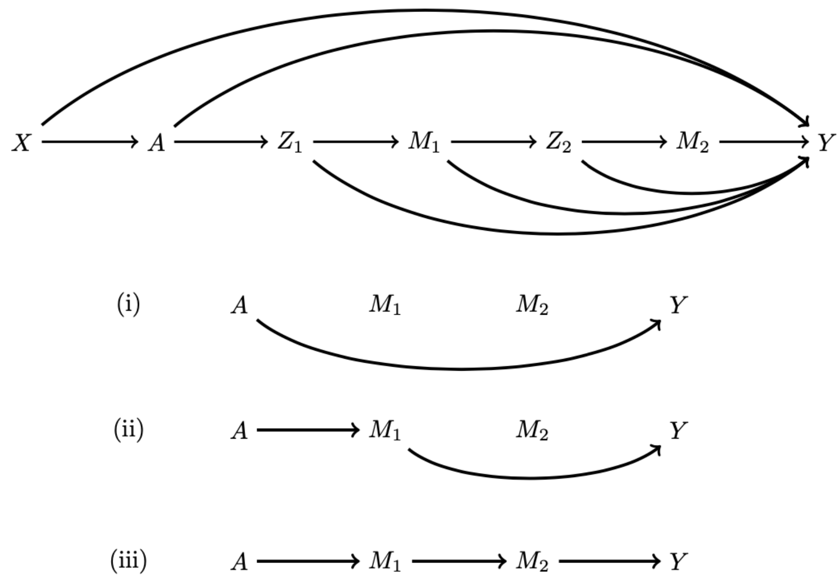

FIG 1. *Causal Relationships with Two Monotonic Mediators Shown in a Directed Acyclic Graph (DAG) and the 3 Monotonic Path Specific Effects (MPSEs). A denotes an initial transition of interest, Y, an outcome, and M_1 and M_2 are two causally ordered, monotonic mediators. The set (X, Z_1, Z_2) captures pre-treatment and intermediate confounders.*

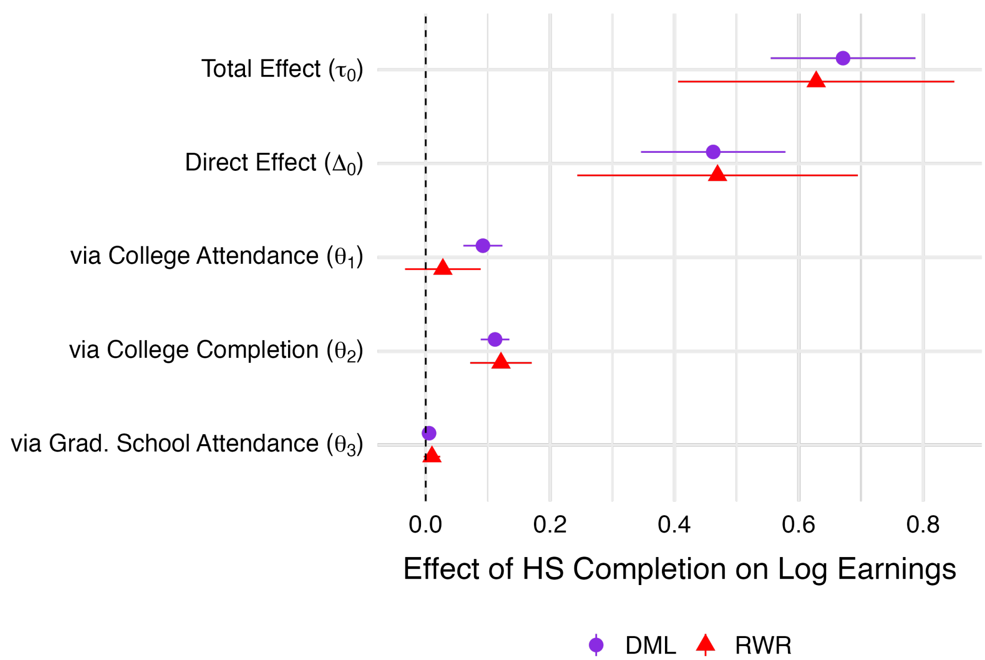

FIG 2. *Decomposition of the Average Total Effect (ATE) of High School Graduation on Logged Earnings via Debiased Machine-Learning (DML) and Regression-With-Residuals (RWR).*

TABLE 1
Direct Effects (Δ_k), Probabilities (π_k) and Covariance Terms (η_k) Involved in Decomposition via Debiased Machine-Learning (DML) and Regression-With-Residuals (RWR).

	Δ_0	Δ_1	Δ_2	Δ_3	π_1	π_2	π_3	η_1	η_2	η_3
DML	0.462	0.200	0.463	0.122	0.427	0.554	0.315	0.006	0.005	-0.016
	(0.059)	(0.034)	(0.046)	(0.029)	(0.009)	(0.015)	(0.022)	(0.007)	(0.007)	(0.009)
RWR	0.469	0.117	0.491	0.160	0.374	0.515	0.219	-0.016	0.071	0.017
	(0.115)	(0.082)	(0.097)	(0.113)	(0.015)	(0.066)	(0.018)	(0.007)	(0.015)	(0.046)

Note: The Δ_k parameters capture the average effect of completing the kth mediator *but no subsequent mediator* on earnings, relative to completing the $k-1$th mediator. For instance, Δ_0 denotes the effect of completing high school (M_1) but not attending college nor, under Assumption 2, completing any subsequent mediators, relative to attending high school but not completing it ($M_0 \equiv A$). The π_k terms capture the average of individuals' counterfactual completion status of the kth mediator under completion of all prior mediators $M_0, \dots M_{k-1}$. For example, π_1 denotes individuals' average counterfactual college attendance, after - possibly contrary to fact - their completion of high school. Finally, the η_k terms refer to the covariance between individuals' own counterfactual completion status of the kth mediator, and their own "gross" effect of completing the kth mediator on earnings. To recall, the "gross" effect of the kth mediator captures the effect of completing that mediator, relative to completing only the $k-1$th mediator, irrespective of whether that effect operates directly (net of subsequent mediators) or via subsequent transitions. For example, η_1 denotes the covariance between each individual's counterfactual college attendance status and their gross effect of college attendance on earnings.

Acknowledgments. Many thanks to Clem Aeppli, Kosuke Imai, Ian Lundberg, Michael Zanger-Tishler, Yi Zhang, and Xiang Zhou, as well as to members of the Harvard C.A.R.E.S. lab and a reviewer from the Alexander and Diviya Magaro Peer Pre-Review program, for helpful feedback and conversations.

SUPPLEMENTARY MATERIAL

A: Parametric, regression-with-residuals (RWR) estimation
Provides details about a parametric, regression-with-residuals estimation procedure.

B: A simulation study
Provides a simulation simulation study of the proposed methods.

C: Sensitivity analysis
Provides a sensitivity analysis for the main empirical results under unobserved confounding.

D: Further details on variable construction and education groups
Provides further information on sample construction and the data used.

E: Results without imputation of missing covariates
Provides empirical results without multiple imputation for missing values.

F: Results under alternative definitions of earnings
Provides empirical results under alternative definitions of the outcome variable.

G: Extension to multivalued, discrete mediators
Provides an extension of the proposed methods to multivalued, discrete mediators.

H: Description of EIFs used in empirical illustration
Provides details on EIFs used in empirical example.

I: Proofs and technical details

REFERENCES

ALBERT, J. M. and NELSON, S. (2011). Generalized causal mediation analysis. *Biometrics* **67** 1028-1038.

ANGRIST, J. D. and CHEN, S. H. (2011). Schooling and the Vietnam-era GI Bill: Evidence from the draft lottery. *American Economic Journal: Applied Economics* **3** 96-118.

ANGRIST, J. D. and KRUEGER, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics* **106** 979-1014.

ANGRIST, J. D. and KRUEGER, A. B. (1992). The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American statistical Association* **87** 328-336.

ASHENFELTER, O. and ZIMMERMAN, D. J. (1997). Estimates of the returns to schooling from sibling data: Fathers, sons, and brothers. *Review of Economics and Statistics* **79** 1-9.

AVIN, C., SHPITSER, I. and PEARL, J. (2005). Identifiability of path-specific effects.

BIRD, K. A., CASTLEMAN, B. L., DENNING, J. T., GOODMAN, J., LAMBERTON, C. and ROSINGER, K. O. (2021). Nudging at scale: Experimental evidence from FAFSA completion campaigns. *Journal of Economic Behavior Organization* **183** 105-128.

BLACK, S. E., DENNING, J. T., DETTLING, L. J., GOODMAN, S. and TURNER, L. J. (2023). Taking it to the limit: Effects of increased student loan availability on attainment, earnings, and financial well-being. *American Economic Review* **113** 3357-3400.

BLEEMER, Z. (2022). Affirmative action, mismatch, and economic mobility after California's Proposition 209. *The Quarterly Journal of Economics* **137** 115-160.

BODORY, H., HUBER, M. and LAFFÉRS, L. (2022). Evaluating (weighted) dynamic treatment effects by double machine learning. *The Econometrics Journal* **25** 628-648.

BRAND, J. E. and XIE, Y. (2010). Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American sociological review* **75** 273-302.

CARD, D. (1994). Earnings, schooling, and ability revisited.

CARD, D. (1999). *The causal effect of education on earnings* **3** 1801-1863. Elsevier.

CARD, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* **69** 1127-1160.

CARNEIRO, P., HECKMAN, J. J. and VYTLACIL, E. J. (2011). Estimating marginal returns to education. *American Economic Review* **101** 2754-2781.

CASTLEMAN, B. L., DEUTSCHLANDER, D. and LOHNER, G. (2020). Pushing college advising forward: Experimental evidence on intensive advising and college success. *EdWorkingPapers. com*.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C. and NEWEY, W. (2017). Double/debiased/neyman machine learning of treatment effects. *American Economic Review* **107** 261-265.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters.

CHETTY, R., DEMING, D. J. and FRIEDMAN, J. N. (2023). Diversifying society's leaders? The causal effects of admission to highly selective private colleges.

COHODES, S. R. and GOODMAN, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics* **6** 251-285.

DANIEL, R. M., STAVOLA, B. L. D., COUSENS, S. N. and VANSTEELANDT, S. (2015). Causal mediation analysis with multiple mediators. *Biometrics* **71** 1-14.

DYNARSKI, S., LIBASSI, C. J., MICHELMORE, K. and OWEN, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review* **111** 1721-1756.

ELLER, C. C. (2023). What Makes a Quality College? Re-examining the Equalizing Potential of Higher Education in the United States.

ELLER, C. C. and DIPRETE, T. A. (2018). The paradox of persistence: Explaining the Black-White gap in bachelor's degree completion. *American Sociological Review* **83** 1171-1214.

FARBMACHER, H., HUBER, M., LAFFÉRS, L., LANGEN, H. and SPINDLER, M. (2022). Causal mediation analysis with double machine learning. *The Econometrics Journal* **25** 277-300.

GOODMAN, J., HURWITZ, M. and SMITH, J. (2017). Access to 4-year public colleges and degree completion. *Journal of Labor Economics* **35** 829-867.

HECKMAN, J. J., HUMPHRIES, J. E. and VERAMENDI, G. (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy* **126** S197-S246.

HOUT, M. (2012). Social and economic returns to college education in the United States. *Annual review of sociology* **38** 379-400.

HURWITZ, M. and HOWELL, J. (2014). Estimating causal impacts of school counselors with regression discontinuity designs. *Journal of Counseling Development* **92** 316-327.

KANE, T. J. and ROUSE, C. E. (1993). Labor market returns to two-and four-year colleges: is a credit a credit and do degrees matter?

KENNEDY, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*.

KENNEY, C. (2004). Cohabiting couple, filing jointly? Resource pooling and US poverty policies. *Family Relations* **53** 237-247.

KIRK, D. S. and SAMPSON, R. J. (2013). Juvenile arrest and collateral educational damage in the transition to adulthood. *Sociology of education* **86** 36-62.

LEWIS, G. and SYRGKANIS, V. (2020). Double/debiased machine learning for dynamic treatment effects via g-estimation. *arXiv preprint arXiv:2002.07285*.

LIN, S.-H. and VANDERWEELE, T. (2017). Interventional approach for path-specific effects. *Journal of Causal Inference* **5** 20150027.

MARE, R. D. (1980). Social background and school continuation decisions. *Journal of the American Statistical Association* **75** 295-305.

MILES, C. H., SHPITSER, I., KANKI, P., MELONI, S. and TCHETGEN, E. J. T. (2017). Quantifying an adherence path-specific effect of antiretroviral therapy in the Nigeria PEPFAR program. *Journal of the American Statistical Association* **112** 1443-1452.

MILES, C. H., SHPITSER, I., KANKI, P., MELONI, S. and TCHETGEN, E. J. T. (2020). On semiparametric estimation of a path-specific effect in the presence of mediator-outcome confounding. *Biometrika* **107** 159-172.

MOUNTJOY, J. (2022). Community colleges and upward mobility. *American Economic Review* **112** 2580-2630.

MOUNTJOY, J. and HICKMAN, B. R. (2021). The returns to college (s): Relative value-added and match effects in higher education.

PEARL, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys* **3**. https://doi.org/10.1214/09-SS057

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical modelling* **7** 1393-1512.

ROTNITZKY, A., ROBINS, J. and BABINO, L. (2017). On the multiply robust estimation of the mean of the g-functional. *arXiv preprint arXiv:1705.08582*.

SCOTT-CLAYTON, J. and WEN, Q. (2019). Estimating returns to college attainment: Comparing survey and state administrative data–based estimates. *Evaluation Review* **43** 266-306.

SMITH, J., GOODMAN, J. and HURWITZ, M. (2020). The economic impact of access to public four-year colleges.

SNYDER, T. D., DE BREY, C. and DILLOW, S. A. (2016). Digest of Education Statistics 2014, 50th Edition. NCES 2016-006.

STEEN, J., LOEYS, T., MOERKERKE, B. and VANSTEELANDT, S. (2017). Flexible mediation analysis with multiple mediators. *American journal of epidemiology* **186** 184-193.

SUGIE, N. F. and TURNEY, K. (2017). Beyond incarceration: Criminal justice contact and mental health. *American Sociological Review* **82** 719-743.

SULLIVAN, Z., CASTLEMAN, B. L. and BETTINGER, E. (2019). College advising at a national scale: Experimental evidence from the CollegePoint initiative.

SWEENEY, M. M. and PHILLIPS, J. A. (2004). Understanding racial differences in marital disruption: Recent trends and explanations. *Journal of Marriage and Family* **66** 639-650.

TURNER, L. J. and GURANTZ, O. (2024). Experimental Estimates of College Coaching on Postsecondary Re-enrollment.

VANDERWEELE, T. J. and VANSTEELANDT, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface* **2** 457-468.

VANDERWEELE, T. J., VANSTEELANDT, S. and ROBINS, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology (Cambridge, Mass.)* **25** 300.

VANSTEELANDT, S. and DANIEL, R. M. (2017). Interventional effects for mediation analysis with multiple mediators. *Epidemiology* **28** 258-265.

VIVIANO, D. and BRADIC, J. (2021). Dynamic covariate balancing: estimating treatment effects over time. *arXiv preprint arXiv:2103.01280*.

WEAVER, V. M. and LERMAN, A. E. (2010). Political consequences of the carceral state. *American Political Science Review* **104** 817-833.

ZHOU, X. (2022a). Semiparametric estimation for causal mediation analysis with multiple causally ordered mediators. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **84** 794-821.

ZHOU, X. (2022b). Attendance, completion, and heterogeneous returns to college: A causal mediation approach. *Sociological Methods Research* 00491241221113876.

ZHOU, X. and PAN, G. (2023). Higher education and the black-white earnings gap. *American Sociological Review* **88** 154-188.

ZIMMERMAN, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics* **32** 711-754.