UCLA

Title

How, and For Whom, Does Higher Education Increase Voting?

Authors

Ahearn, Caitlin E
Brand, Jennie E
Zhou, Xiang

How, and For Whom, Does Higher Education Increase Voting?

Abstract

The college-educated are more likely to vote than are those with less education. Prior research suggests that the effect of college attendance on voting operates directly, by increasing an individual's interest and engagement in politics through social networks or human capital accumulation. College may also increase voting indirectly by leading to degree attainment and increasing socioeconomic status, thus facilitating political participation. However, few studies have empirically tested these direct and indirect pathways or examined how these effects vary across individuals. To bridge this gap, we employ a nonparametric causal mediation analysis to examine the total, direct, and indirect effects of college attendance on voting and how these effects differ across individuals with different propensities of attending college. Using data from the 1979 and 1997 cohorts of National Longitudinal Surveys of Youth, we find large direct effects of college on self-reported voting and comparably smaller indirect effects that operate through degree completion and socioeconomic attainment. We find the largest impact of college on voting for individuals unlikely to attend, a pattern due primarily to heterogeneity in the direct effect of college. Our findings suggest that civic returns to college are not contingent upon degree completion or socioeconomic returns. An exclusive focus on the economic returns to college can mask the broader societal benefits of expanding higher education to disadvantaged youth.

How, and For Whom, Does Higher Education Increase Voting?

Voting is the primary way citizens participate in democratic politics, by electing representatives and shaping policy outcomes. Yet in each election cycle, millions of eligible voters in the United States choose to stay home, with vast implications for representative democracy. Voter turnout is typically 55 to 60 percent for presidential elections, 40 to 45 percent for midterm congressional elections, and generally even lower for local elections (U.S. Bureau of the Census 2018b). Moreover, voting in the United States, and political participation more broadly, is strongly patterned by educational attainment (Leighley and Nagler 2013; Putnam 2001). According to Current Population Survey data, college-educated individuals are 50 percent more likely to vote than those with only a high school diploma (U.S. Bureau of the Census 2018c). Inequality in voter turnout means that the needs and interests of highly-educated citizens may be overrepresented relative to those of less-educated citizens (Bartels 2018; Leighley and Nagler 2013).

Although educational institutions have historically sought to develop civic-minded and engaged citizens (Labaree 1997), some scholars have questioned the causal relationship between college and political participation. They contend that this association may be spurious due to college-goers being more primed for political engagement than non-college-goers (Highton 2009; Kam and Palmer 2008; Tenn 2007). This body of research, however, has generally ignored the possibility that the effects of college may systematically differ across the population. Another line of research, by contrast, suggests that the impact of college on various civic and socioeconomic outcomes may be strongest for those with lower propensities to attend college —

1

i.e., individuals from more disadvantaged backgrounds and with lower academic ability and achievement (Brand 2010; Brand and Xie 2010; Brand, Xu, et al. 2021; Card 2001; Heckman, Humphries, and Veramendi 2018; Skinner and Doyle 2021). If the effect of college on voting follows this pattern of heterogeneity, expanding higher education may lead to greater political engagement among currently underrepresented citizens.

Closely related to heterogeneity in the effects of college are the mechanisms through which those effects operate. Scholars have advanced several theories for how higher education might increase voter turnout. Higher education might directly increase voting through socialization on campus (Bowman 2011; Chatfield and Henderson 2016; Glynn, Huge, and Lunney 2009), the development of civic literacy (Cassel and Lo 1997; Galston 2001; Hillygus 2005), and politically-engaged social networks (Hansen and Tyner 2019; Kingston et al. 2003; Rolfe 2012). Alternatively, a college education may increase voting indirectly, primarily by boosting an individual's SES, thereby increasing access to occupational networks that foster political engagement (Brady, Verba, and Schlozman 1995; Cassel and Lo 1997; Galston 2001; Kingston et al. 2003; Leighley and Nagler 2013). Scholars have also identified bachelor's degree completion as a key pathway through which individuals realize the returns to higher education (Hout 2012). Individuals who attend college may need to complete a degree to realize the civic returns.

Disentangling the contributions of these direct and indirect effects of college on voting provides insight into the relationship between the pecuniary and non-pecuniary benefits of college enrollment. For example, if the voting returns to college depend heavily on post-college socioeconomic attainment, voter turnout should increase most among the subset of college-goers

2

who experience the largest SES returns to higher education. Thus, assessing the degree to which SES mediates the effect of college is essential to understanding not only the extent to which higher education might increase voter turnout but also how it shapes participatory inequality. The proposed direct (Galston 2001; Hillygus 2005) and indirect (Brady et al. 1995; Kingston et al. 2003) pathways may operate more strongly among disadvantaged college-goers. First, colleges' civic missions and liberal education may be particularly consequential for first-generation and lower-income college students, whose social and cultural background may not have as expressly cultivated the commitment to political engagement as that of their more advantaged peers. Higher education may also counter sociopolitical factors and social messaging that suggest to disadvantaged citizens that their votes are inconsequential (Bartels 2018, Wolak 2018). Second, if college enhances the SES of individuals from disadvantaged origins to a greater extent than their more advantaged peers (Brand [forthcoming]; Brand and Xie 2010; Brand, Xu, et al. 2021; Card 2001; Heckman et al. 2018), the former group will likewise benefit more from the indirect effects of college.

To date, few studies have jointly considered the direct and indirect pathways through which college affects voting or explored how those pathways vary across the population. Using a novel approach to causal mediation analysis that can accommodate multiple, causally dependent mediators (Zhou and Yamamoto 2022), this study examines the total, direct, and indirect effects of four-year college attendance on self-reported voting. We also investigate how these effects vary across individuals with different propensities of attending college. Thus, we show not only the degree to which the civic returns to college attendance depend on degree attainment and socioeconomic returns, but also how this dependence differs between individuals with different

3

characteristics. We use data from the National Longitudinal Survey of Youth 1979 and 1997

cohorts. The former allows individuals two decades to realize the socioeconomic returns to

college, thus providing a credible assessment of the mediating role of SES in the effect of college

on self-reported voting. The latter offers a comparison of the total, direct, and indirect effects

through completion for a more recent cohort.

Our study yields several noteworthy findings. First, the direct effects of college

attendance on self-reported voting are larger than the indirect effects through degree completion

or SES attainment. Moreover, both the total and direct effects of college attendance are strongest

for low-propensity individuals, suggesting a pattern of "negative selection." The contribution of

this study is thus three-fold. First, we contribute to the debate on the civic benefits of higher

education by providing strong evidence that four-year college attendance increases political

participation and that the effects are largest among more disadvantaged youth, i.e., those who are

least likely to attend college. Second, using a novel approach to causal mediation analysis, we

advance understanding of the mechanisms by which college increases political participation.

Finally, our results suggest that the civic returns to college are not contingent upon the

socioeconomic returns or even degree completion and that an expansion of higher education has

the potential to narrow the socioeconomic gap in political participation and influence.

BACKGROUND

THE EFFECT OF EDUCATION ON VOTING

Education is one of the most consistent and strongest correlates of voting (Brady et al. 1995;

Hauser 2000; Olsen 1972; Putnam 2001). Yet, some have questioned whether the association is

4

causal, or if college simply serves as a proxy for the effects of social background, pre-college cognitive ability, or psychosocial skills (Berinsky and Lenz 2010; Cassel and Lo 1997; Dee 2004; Highton 2009; Kam and Palmer 2008; Tenn 2007). To address the issue of selection bias, scholars have used covariate adjustment, instrumental variable methods, and machine learning approaches. Many studies find positive causal effects of education on political participation (Dee 2004; Doyle and Skinner 2017; Mayer 2011; Milligan, Moretti, and Oreopoulos 2004; Skinner and Doyle 2021), while others argue that the relationship is largely spurious (Kam and Palmer 2008; Quintelier 2010; Tenn 2007).

To make sense of these ostensibly conflicting findings, we note that most prior research on the effect of college on political participation has focused on average causal effects, without attending to potential effect heterogeneity. The divergence in findings may be reconciled if we examine heterogeneity in the effects of college on political participation. Studies that report positive effects of college on political participation have often used instrumental variable methods that leverage exogenous variation in compulsory schooling laws, child labor laws, or the availability of community colleges (e.g., Dee 2004; Doyle and Skinner 2017; Heckman, Humphries, and Veramendi 2018; Milligan 2004). If there is heterogeneity in the effect of college on voting, then these results apply to students induced into college by the instrument, who are typically relatively disadvantaged and less likely to attend college than the average college-goer. For example, Doyle and Skinner (2017) find that IV estimates of college effects exceed standard OLS estimates for self-reported voting and conclude that civic benefits to higher education appear to accrue largely for individuals on the margins of attendance. Moreover,

5

Skinner and Doyle (2021) find somewhat larger effects of college enrollment on voter registration among those with a lower propensity to enroll.

Additionally, a body of higher education research has found that the effects of four-year college on socioeconomic and civic outcomes differ across the population. In particular, the effects of college appear to vary by selection into college, such that individuals with a lower likelihood of attending and completing college experience different returns from college than those with a higher likelihood (Brand [forthcoming]; Brand 2010; Brand and Xie 2010; Card 2001; Carneiro, Cheng et al. 2021; Heckman, and Vytlacil 2011; Heckman, Humphries, and Veramendi 2018; Zhou and Xie 2020). Some research suggests that the effect of college on civic engagement and self-reported voting is largest for individuals who are unlikely to attend and complete college (Brand 2010; Heckman et al. 2018).

In this study, we contribute to the growing literature on heterogeneous college effects by examining how the effect of college on political participation varies by selection into college. We attend to differential selection bias across subpopulations of interest and assess the sensitivity of our results to different forms of unobserved confounding.

MECHANISMS LINKING COLLEGE AND VOTING

Examining potential mechanisms is critical to understanding the relationship between higher education and voting and why particular subgroups may gain more from college than others. Scholars have suggested both direct and indirect pathways through which four-year college increases voting: directly through socialization and the development of civic literacy and indirectly through socioeconomic attainment and family formation. As an important place of socialization, college can provide access to diverse social experiences, which may contribute to

6

developing a social identity geared toward civic and political participation (Stevens, Armstrong, and Arum 2008). Exposure to racial and ethnic diversity fosters interactions with diverse peers, which can increase civic and political engagement (Bowman 2011). Relatedly, campus social norms that encourage political participation are associated with higher rates of voting (Glynn et al. 2009). The social theory of political participation suggests that more educated citizens are more likely to vote due to more cohesive social networks and stronger social norms around voting (Hansen and Tyner 2019; Rolfe 2012). Social networks developed because of higher education can have lasting effects on political participation into adulthood. Alternatively, negative social experiences on campus, such as racially motivated microaggressions, may spur students to civic action (Baker and Blissett 2018; Skinner and Doyle 2021). The social aspect of college attendance may thus directly encourage individuals toward political participation.

Higher education can also increase civic literacy and subsequent political participation (Cassel and Lo 1997; Chatfield and Henderson 2016; Galston 2001; Hauser 2000; Hillygus 2005; Milligan et al. 2004). Certain types of college learning experiences may have greater impacts on political literacy than others. For instance, civic and political learning is more common in majors that emphasize the political system, social institutions, and inequality (Hillygus 2005; Quintelier 2010) or in liberal arts programs that emphasize general learning (Arum and Roksa 2010). Given that socialization and civic literacy are part of the college experience, and social networks developed during college often extend into adulthood, we characterize these processes as constituting the *direct* effect of college on voting.

The effect of college attendance on voting may also operate through a mediator that is both an outcome of college attendance and a cause of increased political participation. One

7

potential mediator is the attainment of a bachelor's degree. Previous research has found that a

bachelor's degree operates as a gatekeeper to socioeconomic returns to college attendance (Hout

2012), but whether civic returns similarly require degree completion remains uncertain. We thus

distinguish the direct effect of college attendance from the indirect effect via college completion

to assess whether college boosts voting regardless of or because of bachelor's degree attainment.

It is well established that a college education positively affects socioeconomic status

(SES), be it measured by individual earnings, family income, employment, or occupational status

(Hout 2012; Sewell and Hauser 1975). Previous studies suggest that both income (Freeman

2003; Leighley and Nagler 2013; but see Brady et al. [1995] and Wolfinger and Rosenstone

[1980] for some counter-evidence) and occupational position (Freeman 2003; Wolfinger and

Rosenstone 1980) are positively associated with the likelihood of voting. Brady, Verba, and

Schlozman (1995) argue that occupational experiences help workers develop civic skills. The

occupations of college-educated workers tend to cultivate networking, writing, and public

speaking skills, which consequently increase their political engagement. College graduates may

also experience greater encouragement to participate in political activities, possibly because they

are perceived as having greater access to socioeconomic resources or more influential networks

(Hauser 2000). Kingston and colleagues (2003) contend that SES "is critical in shaping interests,

outlooks, and social behavior," rendering education "socially consequential only indirectly

through its impact on economic status" (p. 57).

In these cases, greater socioeconomic attainment is understood to increase access to and

social encouragement for political activities, including voting. A corollary of this process is that

lower-SES citizens face socioeconomically-driven obstacles that reduce their likelihood of

8

voting. Such obstacles might be a lack of reliable transportation to the polls (U.S. Bureau of the

Census 2012), frequent moving that prevents registration (Phinney 2013; Wolfinger and

Rosenstone 1980), or other adversities that demand their time and energy. Thus, college may

increase voting indirectly through SES not only by increasing the proclivity to vote but also by

reducing socioeconomic barriers to political engagement.

Education also has important effects on family outcomes, including increased likelihood

of marriage (Schneider 2011), reduced likelihood of single parenthood (Lundberg, Pollak, and

Stearns 2016), and increased marital stability (Amato 2010). Married individuals consistently

exhibit higher rates of political participation than others (Leighley and Nagler 2013). College

may thus also boost voting through its impact on family formation and stability. Moreover,

because family formation and stability itself facilitate socioeconomic attainment (e.g., Killewald

and Gough 2013; Smock, Manning, and Gupta 1999), the mediating effect of family may further

operate through SES, constituting a causal chain characterized by education → family → SES →

voting.

HETEROGENEITY IN THE MECHANISMS LINKING COLLEGE AND VOTING

As with the total effect of college on voting, there is little reason to assume that the mediating

effects are constant across the population. While no prior research (to our knowledge) has

considered potential heterogeneity in the mediating pathways between college and voting, prior

work on heterogeneous economic returns to college offers some directions to consider. In

particular, the effects of four-year college attendance and completion on earnings appear to be

high for individuals who are less likely to attend and complete college (Brand and Xie 2010;

Cheng et al. 2021; Card 2001). Relatedly, recent research suggests that college circumvents

9

socioeconomic disadvantage for those on the margin of school continuation (Brand [forthcoming]; Brand, Xu, et al. 2021; Heckman et al. 2018). From this perspective, we would expect the indirect effect of college via SES to be greater for individuals who are unlikely to attend college. The direct effects of college on voting may similarly vary. Students from more disadvantaged backgrounds or from racially minoritized groups, who are more commonly on the margin of four-year college attendance, may have social encounters on college campuses that encourage social activism (Baker and Blissett 2018; Jack 2019). Likewise, more disadvantaged college-goers may face lower pre-college social pressure to vote, and thus may face greater increases in positive social pressure from college attendance.

Scholars have shown that the effects of college attendance on marriage and marital stability vary with the propensity to complete college. In this case, however, the largest effects accrue to those who are most likely to complete college (Musick et al. 2012). On the other hand, the effect of college on reducing fertility is larger among more disadvantaged college-goers (Brand and Davis 2011). We thus remain agnostic with respect to how the mediating effect of family characteristics varies across population subgroups. We nevertheless include variables related to family formation and marital stability in our models given the important effects of college on family and of family on political participation.

Despite the well-documented association between education and political participation and theories on different mechanisms leading to increased voting, the mediating pathways between college and voting remain poorly understood. Prior studies have not explored different pathways using a causal mediation framework. In addition, no previous research has examined how college's direct and indirect effects on political participation vary across the population. If

10

college increases voting more strongly for disadvantaged youth, as recent research suggests, studies should focus on mechanisms explaining this important source of heterogeneity. In this paper, we investigate not only the extent to which degree attainment, family characteristics, and socioeconomic status mediate the effect of college on self-reported voting but also how the direct and indirect effects vary by the propensity of attending college.

VOTER REPORTING AND REPRESENTATION BIAS IN SURVEY DATA

Voting rates in survey data tend to be considerably higher than actual turnout rates (Berent, Krosnick, and Lupia 2016; Debell et al. 2020). Scholars have argued that two processes bias survey voting rates: overrepresentation of voters among respondents and vote overreporting among non-voters (Ansolabehere and Hersh 2012; Enamorado and Imai 2019; Goldberg and Sciarini 2019; Selb and Munzert 2013). Some validation studies, which use official data to assess the accuracy of self-reported voting behavior, have identified high rates of voter overreporting in surveys and have shown that overreporting tends to increase with education (Ansolabehere and Hersh 2012; Enamorado and Imai 2019). This suggests that analyses of survey data, such as those in our study, may overstate the effect of education on voting.

On the other hand, the source of the survey turnout gap in this study, and thus the implications for our results, remains unclear. First, validation studies in the U.S. have generally relied on political surveys, such as the American National Election Study (ANES), which draw more politically motivated respondents than a general survey like the NLSY. A less political sample may feel less pressure to misreport their voting behavior, resulting in a smaller turnout gap (Debell et al. 2020). Second, validated data of U.S. voters are imperfect. A recent report by the Pew Research Center (2018) shows that commercial voting data sets tend to be incomplete

11

when examined individually. Even when combined, these data sets are more likely to miss disengaged, young, Hispanic, and low-income individuals (Pew 2018). In an analysis of voter validation methods, Berent, Krosnick, and Lupia (2016) similarly find that "the process of matching government records often failed to locate records of respondents who were truly registered and had voted" (p 615). They argue that matching survey respondents to government records underestimates the proportion of Americans registered to vote. Among respondents whose self-reports can be validated against government records, the accuracy of self-reports is extremely high. Their findings imply that while validated turnout estimates appear to be more accurate than self-reports because they produce lower turnout estimates, "the apparent accuracy is most likely an illusion" (Berent et al. 2016, p. 599).

Moreover, and perhaps more importantly, our primary aim in this study is to assess heterogeneity in the effect of college on voting by the propensity to enroll in college. We find that the effect of college is larger for more marginal college-goers, who tend to be more disadvantaged and lower-achieving. Bias-induced invalidation of these effects would require either more voting overreporting or greater overrepresentation of voters among lower propensity college-goers than their higher propensity peers. To our knowledge, there is no evidence from existing literature that either of these processes is at play.

DATA AND MEASURES

We use data from the 1979 and 1997 National Longitudinal Surveys of Youth (NLSY) cohorts, which follow respondents from adolescence through adulthood. Both surveys interviewed respondents annually or biennially, collecting data on sociodemographic background, academic achievement, attitudes, educational attainment, and family and labor market outcomes during

12

adulthood. Our main analyses use the 1979-2006 waves of the NLSY79 cohort, which began

with adolescents aged 14-22 in 1979. We restrict our NLSY79 sample to 14-17-year-old

respondents at the baseline survey in 1979 ($n = 5,582$) who had completed at least the 12th grade

($n = 4,548$). We set these sample restrictions to ensure that all variables used to predict college

are measured before college, particularly academic ability, and to compare college-goers to those

who completed only a high school education. We also focus on the region of common support in

terms of the estimated propensity scores, as described below ($n = 4,085$). We further restrict the

sample to respondents who answered the question about voting in the 2006 general election ($n = 2,961$). This reduction in sample size largely reflects attrition over 27 years of follow-up.

In 2006, the NLSY79 asked respondents whether they voted in the 2006 general election.

NLSY political questions are modeled from questions from the American National Election

Studies (ANES). Self-reported voting is based on a question that asks respondents which

statement best describes them: (1) I am sure I voted; (2) I usually vote, but didn't in 2006; (3) I

thought about voting in 2006, but didn't; and (4) I did not vote in the November 2006 election.

Some respondents refused to answer or reported that they didn't know. We code those who

stated that they are sure they voted as voting and all other non-missing categories as did not

vote.[1] Among NLSY79 respondents who have at least a high school diploma, 69 percent report

voting.

---

[1] The NLSY voting measure benefits from four factors. First, misreporting may be lower in reporting turnout for midterm than for presidential elections. Second, as noted above, misreporting may be lower in non-political surveys like the NLSY relative to political surveys like the American National Election Studies. Third, accuracy is greater in face-to-face interviews, like the NLSY, than in phone interviews (DeBell et al. 2020). Fourth, the NLSY question wording is intended to reduce social desirability pressures by allowing nonvoters to express their identity as habitual voters (option 2) and to indicate that they voted only if they were sure.

13

As noted, people who do not vote are disproportionately unlikely to participate in surveys. To assess the robustness of our results to potential bias due to the overrepresentation of voters among those who responded to the voting question in NLSY, we have conducted a sensitivity analysis where individuals who were not interviewed for or did not answer the voting question were coded as non-voters. The findings from this analysis are discussed briefly in the Results section and in greater detail in Appendix D.[2]

We define the treatment variable as whether a respondent attended a four-year college by age 20.[3] Pre-college covariates include individual, school, and family characteristics that are known to predict college attendance. Our sociodemographic characteristics include race, gender, family socioeconomic status, family structure, religion, and region. We also use measures of ability (i.e., the 1980 Armed Services Vocational Aptitude Battery, adjusted for age and standardized following Brand and Xie [2010]), psychosocial characteristics, college preparation and expectations, delinquent activity, high school characteristics, and family formation at age 18.

We examine three pathways through which college attendance may increase voting: bachelor's degree completion by age 25, family formation and stability during 1990-1998, and socioeconomic attainment during 1998-2006. By temporally ordering these mediators, we

[2] If we code all those who did not respond as non-voters, 53 percent report voting. This estimate is slightly below that of Current Population Survey (CPS) for the same education and age groups (i.e., 57 percent, authors' calculations). The CPS voting measure also codes missing as non-voters. However, the CPS is a cross-sectional survey and thus those missing are not missing due to sample attrition, as is true when we code survey non-response as non-voters in the NLSY. The voting gap between college and non-college graduates is roughly the same (i.e., about 20 percentage points). Thus, even if the sample is under-representing non-voters, within-sample analyses may still be reasonable if we believe that the CPS is less susceptible to voter overrepresentation and overreporting.
[3] Examining traditional-aged college students allows sufficient time for respondents to realize their post-college socioeconomic status and family formation. However, we acknowledge that this restriction misses some non-traditional aged college students who are more socioeconomically disadvantaged and less likely to vote. To examine whether these college-goers experience similar effects of college on voting, we ran additional analyses examining the effect of college attendance by age 25, and the mediating effect of bachelor's degree completion by age 30. Results from this analysis are nearly identical to our main analyses.

14

disentangle the unique causal paths via college completion, family characteristics, and socioeconomic attainment successively. We use three variables to gauge family formation and stability: the proportion of time respondents were married, the proportion of time they were unmarried parents, and the number of family transitions between 1990-1998. To assess socioeconomic attainment, we use a battery of variables reflecting employment and wages, job stability, poverty and social assistance, and residential turnover and homeownership between 1998-2006. These include the proportion of time out of the labor force, average wages, number of jobs held, job tenure at age 40, the proportion of time in poverty, cumulative welfare received, number of residential moves between 2000 and 2006, and homeownership status in 2004.

We present descriptive statistics of the political participation outcome, all pre-college covariates, and all family and socioeconomic mediators in Table 1. Our descriptive statistics of pre-college covariates are consistent with well-documented socioeconomic and achievement differences between college- and non-college-goers. Except for a few covariates, these differences are statistically significant at the $p < .05$ level. Descriptive statistics of the post-treatment mediators also align with expected differences by college attendance. College-goers have more stable family lives and higher socioeconomic status. As expected, college-goers are more likely to report voting in the 2006 general election than individuals who did not attend college. These differences are also statistically significant at the $p < .05$ level.

The mediating effects of family and socioeconomic attainment are realized many years after college attendance. Using the NLSY79 allows sufficient time for these life events to unfold and subsequently influence voting during middle adulthood. However, the experiences of this older cohort, who attended college in the 1980s, may be less applicable to current college

15

students. For this reason, we supplement our main findings with analyses from the 1997-2010

waves of the NLSY97 cohort, which began with adolescents aged 12-17 in 1997. Our outcome is

whether respondents voted in the 2010 national election, the only post-college voting item asked

of the full NLSY97 sample. Among NLSY97 respondents who have at least a high school

diploma, 40 percent report voting in the 2010 election. We use the same approach to construct

this measure as for the NLSY79 data. Due to the temporal proximity of the 2010 voting outcome

to this cohort's college completion, we cannot examine the mediating effects of family

characteristics and socioeconomic attainment using the NLSY97. Thus, our supplemental

analyses focus on estimating heterogeneity in the total effect of college attendance and the

mediating effect of college completion on voting. Pre-college covariates used in these analyses,

and differences in those covariates by college attendance status, are like those for the NLSY79

cohort. We report descriptive statistics for the NLSY97 variables in Appendix A.

ANALYTIC STRATEGY

ESTIMATING PATH-SPECIFIC CAUSAL EFFECTS

In this study, we aim to disentangle the specific pathways through which college affects political

participation. This goal poses a methodological challenge because existing methods for causal

mediation analysis have largely focused on the role of a single mediator or of a set of mediators

considered as a whole (Baron and Kenny 1986; Imai, Keele, and Yamamoto 2010; VanderWeele

2015). In the presence of multiple mediators, it is often assumed that these mediators are causally

independent, i.e., they do not affect each other. This assumption, however, is strong, untestable,

and unrealistic in our application. It is well documented that college completion, family

16

formation, and marital stability influence individuals' socioeconomic status (Killewald and

Gough 2013; Smock et al. 1999).

To accommodate the reality that our three mediators are causally dependent, we analyze

the causal diagram shown in the upper panel of Figure 1. This diagram consists of five sets of

temporally ordered and causally dependent variables from the NLSY79: (1) four-year college

attendance by age 20; (2) four-year college completion by age 25; (3) variables reflecting family

characteristics from about age 25 to the mid-30s (i.e., in 1990-1998); (4) variables reflecting SES

from the mid-30s to early 40s (i.e., in 1998-2006); and (5) voting participation in the early 40s

(i.e., in 2006). In this diagram, the baseline covariates that we use to adjust for confounding are

omitted. The lower four panels of Figure 1 represent four possible mechanisms through which

college attendance increases voting: (a) college attendance may increase voting directly; (b)

college attendance may influence college completion, which in turn affects voting either directly

or via family characteristics and/or SES; (c) regardless of college completion status, college

attendance may influence family formation and stability, which in turn affects voting either

directly or through increased SES; (d) regardless of college completion status and family

characteristics, college attendance may increase SES, which in turn leads to higher levels of

voting.

Following Pearl (2009), we interpret a causal diagram as a nonparametric structural

equation model with independent errors. This model assumes no unobserved confounding exists

for any treatment-mediator, treatment-outcome, mediator-mediator, or mediator-outcome

relationships. In our context, it implies that conditional on the precollege covariates described in

the previous section, (a) no unobserved variables affect both college attendance and any of the

17

college completion, family, SES, and voting outcomes; (b) no unobserved variables affect both

college completion and any of the family, SES, and voting outcomes; (c) no unobserved

variables affect both family characteristics and any of the SES and voting outcomes; and (d) no

unobserved variables affect both SES and voting. Recognizing that these assumptions are strong

and untestable, we conduct a series of sensitivity analyses that investigates the direction and

magnitude of potential bias when some of these assumptions are violated. These analyses are

described briefly in the Results section and in more detail in Appendix D.

Under the above assumptions, the path-specific effects corresponding to mechanisms (a),

(b), (c), and (d) in Figure 1 are nonparametrically identified (Avin, Shpitser, and Pearl 2005;

VanderWeele 2015; Zhou 2022), i.e., expressed in terms of observed data without any functional

form assumption. Specifically, if we let A denote treatment (i.e., college attendance), Y the

outcome of interest (i.e., voted in 2006), X the vector of pretreatment covariates, C the college

completion mediator, L the set of mediators reflecting family formation and stability, and M the

set of mediators reflecting SES, the average total effect (ATE) of college on voting can be

decomposed into four components:

$$ATE = \tau_{A \rightarrow Y} + \tau_{A \rightarrow CY} + \tau_{A \rightarrow LY} + \tau_{A \rightarrow M \rightarrow Y}, (1)$$

where the quantities $\tau_{A \rightarrow Y}$, $\tau_{A \rightarrow CY}$, $\tau_{A \rightarrow LY}$ and $\tau_{A \rightarrow M \rightarrow Y}$ reflect panels (a), (b), (c), and (d) in Figure

1, respectively. Their detailed expressions and further explanation are given in Appendix B.

To estimate these path-specific effects, we use an imputation estimator proposed in Zhou

and Yamamoto (2022), which involves fitting four outcome models for voting participation. The

first is conditional on college attendance and the baseline covariates only (i.e., a model for

$E[Y \vee X, A]$). The second is conditional on college attendance, the baseline covariates, and

18

college completion (i.e., a model for $E[Y\vee X,A,C]$). The third is conditional on the college attendance, baseline covariates, college completion, and the mediators reflecting family formation and stability in 1990-98 (i.e., a model for $E[Y\vee X,A,C,L]$). The fourth is conditional on college attendance, the baseline covariates, college completion, the mediators reflecting family formation and stability in 1990-98, and the mediators reflecting SES in 1998-2006 (i.e., a model for $E[Y\vee X,A,C,L,M]$). After fitting these outcome models, we fit three additional models for the conditional expectations: $E[E[X,A=1,C]\vee X,A]$, $E[E[X,A=1,C,L]\vee X,A]$ and $E[E[X,A=1,C,L,M]\vee X,A]$, which are then used to evaluate the path-specific effects in equation (1).

We use boosted regression trees (Friedman 2001), a flexible machine learning method with strong predictive accuracy, to fit the above models. We use the R package *gbm* (Ridgeway 2019) to construct the regression trees, allowing for two-way interactions and using a learning rate of 0.02. The numbers of trees used for these outcome models are selected via five-fold cross-validation. Standard errors and confidence intervals for the path-specific effects are estimated via the nonparametric bootstrap.

EVALUATING TREATMENT EFFECT HETEROGENEITY

In addition to assessing the average total effect of college and its direct and indirect components, we also investigate how these effects vary across individuals. Following previous research on heterogeneous effects of college on socioeconomic and civic outcomes (e.g., Brand 2010; Brand and Xie 2010; Cheng et al. 2021), we use the propensity score of attending college as a summary index of pre-college advantage, i.e., the conditional probability of attending college given the precollege covariates described in the previous section. As a robustness check, we also examined

19

heterogeneity by two alternative indicators of socioeconomic and academic advantage: parental income and measured ability.

We fit the propensity score models for college attendance using an iterative procedure outlined by Imbens and Rubin (2015), which leads to a fairly flexible specification with good balancing properties. After beginning with a set of key baseline covariates, we considered additional possible covariates in turn. We identified covariates, including higher-order and interaction terms, that produce a likelihood ratio statistic that exceeds a pre-set constant when added to the model. This procedure involved roughly 5,800 regressions and resulted in a model with 23 linear terms, one higher-order term, and 19 interaction terms. The final set of covariates identified through this process is described above. We further eliminated units with no common support, i.e., treated cases with values higher than the highest propensity score among the controls ($p(X) = 0.921$; $n = 13$) and untreated cases with values lower than the lowest propensity score among the treated ($p(X) = 0.002$; $n = 411$).

To assess the heterogeneity of the direct and path-specific effects described previously, we extract the "individual-specific" treatment effect estimates inside the expectation operators derived from the expressions of the components of equation 1 (see Appendix B). Then, for each of these individual-specific estimates of $\tau_{A \to Y}$, $\tau_{A \to CY}$, $\tau_{A \to LY}$, and $\tau_{A \to M \to Y}$, we fit a regression model for their conditional mean given the estimated propensity score (in auxiliary analyses, parental income and cognitive ability). To allow for sufficient nonlinearity, we use a smoothing spline with three degrees of freedom (Hastie 2017). Standard errors and confidence intervals for these conditional means are estimated via the nonparametric bootstrap.

20

RESULTS

TOTAL AND PATH-SPECIFIC EFFECTS OF COLLEGE ATTENDANCE ON VOTING

We begin by examining the total and path-specific effects of four-year college attendance on

voting for the overall sample. Results are presented in Figure 2. The estimated effects in these

figures reflect the percentage-point increase in the rate of self-reported voting associated with

attending college. Figure 2 indicates that, overall, college attendance increases self-reported

voting rates by about 12 percentage points. The estimated direct effect of college attendance is

slightly less than two-thirds of the total effect (7.5 percentage points), and the estimated

combined mediating effect of bachelor's degree completion, family formation, and

socioeconomic attainment accounts for the remaining 4.5 percentage points.[4] This suggests that

the effect of college attendance on voting operates primarily directly, whereas the mediating

effects of degree completion, family, and SES are comparatively weaker (though still

significant).

HETEROGENEOUS EFFECTS OF COLLEGE ATTENDANCE ON VOTING

We next assess heterogeneity by the propensity for four-year college attendance in the effect of

college attendance on self-reported voting. Larger effects of college on voting for those with

higher propensity scores would suggest that greater increases in self-reported voting accrue to

those more likely to attend college; larger effects for those with lower propensity scores, on the

---

[4] We acknowledge the possibility that enrollment in a two-year college also increases voting. To examine whether
analyses are sensitive to expanding our treatment to include two-year attendance, we estimated the effect of two- or
four-year attendance by age 25, and the mediating effect of bachelor's degree completion by age 30. The estimated
total effect from this analysis is somewhat smaller, and the estimated direct effect is about half of the total effect.
This suggests that the effect of two-year college attendance on voting are smaller than the effect of four-year college
attendance.

21

other hand, would suggest that those unlikely to attend college gain the most. We present our results for the total, direct, and indirect effects of college attendance in Figures 3 and 4. In the left panel of Figure 3, the dotted and dashed lines show that the probability of voting increases with the propensity of attending college regardless of attendance status. However, the probability of voting is particularly low for high-school graduates with a low propensity to attend college. As a result, the effect of college attendance on voting is higher among low-propensity than among high-propensity individuals, for whom the probability of voting is relatively high whether or not they attended college.[5] The right panel depicts how the effect of college attendance, estimated as the difference between the estimated probability of self-reported voting with and without college attendance, varies by the propensity score. The total estimated effect of attendance on voting for respondents with the lowest propensity to attend is about 17 percentage points and decreases to about 6 percentage points for those with the highest propensity scores. This heterogeneity pattern is consistent with Heckman et al.'s (2018) results using instrumental variable models.

Figure 4 shows that the direct effect of college attendance on voting follows a similar pattern. Individuals who are least likely to attend college experience larger direct effects of college attendance on voting (about 12 percentage points) than those who are most likely to attend college (about 3 percentage points). We do not, however, find much heterogeneity in the indirect effects via degree completion, family, or SES. Thus, the strong pattern of negative selection in the total effect of college attendance on voting is primarily due to negative selection in the direct effect of college. The effect of college attendance on voting appears to operate

[5] Analyses that estimate the effect of any two- or four-year college attendance on voting produce a similar pattern of heterogeneity.

22

largely through pathways other than degree completion, family formation, or socioeconomic attainment. These results are consistent with the hypothesis that college education increases political participation directly through socialization and the development of civic literacy during college, especially for individuals from socioeconomically disadvantaged backgrounds.

To assess the robustness of our estimates of heterogeneity in the total and path-specific effects by the propensity score, we also estimated heterogeneity in total, direct, and mediating effects by two key covariates that shape selection into college: parental income and measured cognitive ability. These results are presented in Appendix Figures C.1 and C.2, respectively. Individuals with higher parental income or higher measured ability have lower total and direct effects of college attendance on voting than individuals with lower parental income or measured ability scores. Estimates of heterogeneity in the total, direct, and indirect effects of college attendance by parental income and cognitive ability align with the effects by the propensity of college.

SUPPLEMENTARY ANALYSES ON HETEROGENEOUS EFFECTS OF COLLEGE ON VOTING

The NLSY79 respondents were born in the early 1960s. While this gives us sufficient time to assess the mediating effects of post-college experiences on political participation, the experiences of this cohort may not well characterize those of more recent cohorts. Thus, we conduct a parallel analysis using the NLSY97 cohort, who were born in the 1980s, and who were asked about their participation in the 2010 general election. As noted earlier, the timing of the 2010 election does not allow us to assess the mediating effects of family and SES for this cohort. We thus restrict this supplementary analysis to assessing heterogeneity in the total effect of college attendance and the mediating effect via college completion. We show the results of the

23

total and mediating effects in Figures 5 and 6. Overall self-reported voting rates are lower in this sample (see appendix Table A.1), presumably because the NLSY97 respondents were much younger (i.e., in their late 20s to early 30s) in 2010 than the NLSY79 respondents were in 2006 (i.e., in their early 40s), and voter turnout tends to increase with age (U.S. Bureau of the Census 2018a). Still, results from these analyses largely align with results from the NLSY79 cohort. The total effect of attendance on voting (see Figure 5) is about 13 percentage points, and the direct effect comprises about two-thirds of that effect (9 percentage points). Our results on effect heterogeneity for the NLSY97 cohort (see Figure 6) are consistent with those reported in Figures 3 and 4. The total and direct effects of college attendance, but not the mediating effect via bachelor's degree completion, decrease as the propensity of college attendance increases.

SENSITIVITY ANALYSES

We conducted two sets of sensitivity analyses to address potential biases in our results. First, to address the impact of voter overrepresentation in our sample, we replicated our main NLSY79 and NLSY97 analyses with an imputed measure of voting. Rather than omitting individuals who did not answer the voting item, we impute all individuals who were missing data for the voting question (either due to missing in the wave or missing the specific item) as non-voters. This assumption is grounded in previous research showing that individuals who do not participate in surveys tend to vote at lower rates than those who participate (Goldberg and Sciarini 2019; Lahtinen et al. 2019). Still, this is a very strong assumption and is likely to overrepresent non-voters.

As with our main analyses, these results suggest a pattern of negative selection in the direct effects of four-year college attendance on voting for both cohorts. However, they indicate

24

that if the NLSY respondents on the voting item were much more likely to vote than those who were missing on that item, our results might have underestimated the mediating effects of college completion and socioeconomic attainment while overestimating the direct effects of college attendance. These auxiliary analyses do not address the issue of potential overrepresentation of voters in the initial survey, as they are conducted on a sample that remained in the survey long enough to provide information on college-going. Nevertheless, they help us understand the potential consequences of voter overrepresentation in the follow-up surveys. Even assuming this very high level of voter overrepresentation bias, our primary finding that low-propensity college-goers experience a greater increase in voting due to a larger direct effect of college attendance remains robust. See Appendix D for the full results of these analyses.

Second, our causal mediation analyses rest on the strong and untestable assumption that no unobserved confounding exists for any of the causal relationships represented in Figure 1. Although we have adjusted for an array of baseline covariates in our analyses, there may still be unobserved individual attributes that affect both college attendance and political participation, or that affect both the mediators and political participation. We therefore conducted sensitivity analyses that explore the degree to which our estimates of total and direct effects are robust to the presence of unobserved confounding of the treatment-outcome and mediator-outcome relationships. Specifically, we use the bias factor approach developed by VanderWeele (2010) and VanderWeele and Arah (2011) (see Brand et al. [2019] for a recent application).

We begin by assessing the impact of an unobserved confounder of the treatment-outcome relationship. We show that if an unobserved trait increases voting by 10 percentage points and differs across college attendance status by 20 percentage points, the bias-adjusted estimate of the

25

total effect will still be statistically significant and substantively similar to the original estimate. Next, we assess the impact of bias due to an unobserved confounder of the mediator-outcome relationship. For example, neighborhood-level social capital, which is not observed in this study, may affect both the mediators and political participation. Results from this analysis suggest that our main conclusion of a substantial direct effect of college attendance on voting is fairly robust to such potential confounding. We further show that, given large propensity score differences in the effect of college attendance on voting (as high as 11 percentage points), our finding of negative selection in the total and direct effects of college attendance is unlikely to be explained away by differential selection bias. Additional details and full results of these analyses are presented in Appendix D.

DISCUSSION AND CONCLUSION

In the U.S., more-educated and higher-SES citizens persistently vote more than less-educated and lower-SES citizens (Leighley and Nagler 2013). In this study, we asked whether four-year college attendance might reduce inequality in voter turnout by increasing voting rates among those from disadvantaged backgrounds and how post-college socioeconomic attainment contributes to this process. Using a machine-learning-assisted causal mediation analysis, we examined how bachelor's degree completion, family formation and stability, and various measures of socioeconomic attainment mediate the effect of college on voting, and the extent to which the total, direct, and indirect effects of college vary across individuals with different likelihoods of college attendance.

26

Our analyses of the NLSY79 cohort yield three notable findings. First, the estimated direct effect of college on self-reported voting is large, constituting about two-thirds of the total effect of attendance. If our models fully account for selection into college and selection into the mediators, this finding suggests that the college experience directly enhances voter turnout. Second, degree completion, post-college family characteristics, and socioeconomic attainment mediate the effect of college on self-reported voting, but the impact is smaller than the direct effect. Third, the total effect of college on voting exhibits a strong pattern of negative selection. Individuals with a lower likelihood of attending college, who tend to have more disadvantaged backgrounds, experience greater increases in self-reported voting due to college attendance than individuals from more privileged backgrounds. We find that the negative selection in the total effect is primarily due to a strong direct effect of college among low-propensity individuals. Our results suggest that self-reported voting increases most among individuals who are least likely to attend college not because they have the largest socioeconomic returns, but because they benefit more from the opportunities provided and skills developed during college than their more advantaged peers.

With a rich set of pre- and post-college variables and its longitudinal structure, the NLSY79 enables us to assess the long-term effects of college and how they unfold over the life course. Nevertheless, the NLSY79 cohort attended college during the 1980s, potentially reducing the direct relevance of our results to more current effects of higher education. To address this challenge, we ran supplementary analyses on the more recent NLSY97 cohort, who attended college in the early 2000s. Due to the young age of the NLSY97 survey respondents, we cannot examine the mediating pathways via family and socioeconomic attainment. Still, these results

27

support our finding of negative selection in the direct effect of college attendance on voting and provide evidence that this effect appears relatively early in adulthood.

Taken together, our findings suggest that the expansion of higher education may reduce inequality in voter turnout, even when individuals do not complete a four-year degree or attain high socioeconomic status. Despite extensive speculation on the processes through which college affects voting, prior work has not assessed the mediating role of post-college socioeconomic attainment. Thus, our study makes important inroads into identifying the potential civic benefits of expanding higher education to relatively disadvantaged groups and the conditions upon which those benefits depend. Scholars should continue to examine the mediating role of socioeconomic attainment in college's additional civic consequences, such as other forms of political participation, volunteering, and pro-social attitudes, and how the direct and mediating effects vary across the population.

At least three potential sources of bias may have affected our results. First, some studies have suggested that vote overreporting among more educated respondents biases the estimates of college effects on voting (Ansolabehere and Hersh 2012; Enamorado and Imai 2019). However, more recent studies using high-quality data have found that the survey turnout gap is driven more by the overrepresentation of voters than by overreporting (Lahtinen et al. 2019; Goldberg and Sciarini 2019). This recent work suggests that upward bias in our estimates of college attendance on self-reported voting is likely limited. Moreover, our focus is on heterogeneity in the effect of college. Previous research has not provided evidence for greater bias among low-propensity compared to high-propensity college-goers. Without data on the actual voting behavior of respondents and non-respondents, we cannot ascertain the level of bias in our estimates.

28

Sensitivity analyses suggest that if there was a significant overrepresentation of voters in the sample who responded to our outcome variables, our results might have overstated the total and direct effects of college attendance on voting. However, the finding of negative selection in the direct effect remains robust.

Second, our analyses rest on the strong and untestable ignorability assumption, i.e., that no unobserved confounding exists for the total, direct, and indirect effects of college on voting. Because we cannot expect to observe and adjust for all possible confounders of the total and mediating effects of college on political participation, we ran a series of sensitivity analyses to understand how our estimates would respond to unmeasured factors (see Appendix D). The results suggest minimal unobserved confounding. We show that failing to adjust for an unobserved confounder in the treatment-outcome relationship or the mediator-outcome relationship is unlikely to meaningfully impact our estimates. Third, our estimates of mediating effects may be underestimated due to measurement error of family characteristics or socioeconomic attainment. Such measurement error would result in an upward bias in the direct effect of college on voting (Vanderweele, Valeri, and Ogburn 2012). However, our analyses include multiple measures for each of our mediators and use adaptive machine learning models, thus reducing the likelihood that our results suffer from substantial bias due to measurement error.

While our analyses provide strong support for a direct effect of four-year college on voting, we cannot adjudicate between specific explanations for this effect. Some have suggested that this direct effect stems from the college experience, including socialization on campus (Bowman 2011; Chatfield and Henderson 2016; Glynn et al. 2009), the development of civic and

29

political literacy during college (Cassel and Lo 1997; Galston 2001; Hillygus 2005), or pressure to conform to their more politically active social networks' norms and expectations (Hansen and Tyner 2019; Kingston et al. 2003; Rolfe 2012). Given that we find stronger direct effects for lower-propensity college-goers, future studies in this area would benefit from assessing which direct processes are particularly salient for this population. For example, the change in social pressure to vote may be largest for relatively disadvantaged college-goers, or the increase in voting may reflect a response to social interactions that incite activism (Baker and Blissett 2018). Although we cannot distinguish these explanations, our results suggest that the direct effects of college on voting, and ostensibly the social norms and networks that maintain them, are conceptually distinct from various measures of family formation and socioeconomic attainment.

While we examine heterogeneity by individual characteristics, future studies may additionally examine the role of different college experiences. For example, there is some evidence that online coursework is associated with reduced civic engagement (Chatfield and Henderson 2016). ~~This suggests that~~If limited campus time reduces the civic impact of college, the effects of college on voting may be similarly smaller for commuter students or those who work full time due to less exposure to ~~non-classroom~~campus experiences. Since relatively disadvantaged college-goers are more likely to commute (Schudde 2011), an examination of commuter students may provide additional insight into differential effects within subpopulations of low-propensity college-goers and the mechanisms that lead to civic engagement. Finally, while we focus on four-year college attendance and bachelor's degree completion, recent research has suggested that increased proximity to a community college leads to increased voting

(Doyle and Skinner 2017). Future research in this area may investigate whether the effects identified in the present study extend to alternative post-secondary pathways and degrees.

In the early years of American higher education, colleges were charged with developing an educated citizenry to maintain a vibrant democracy. Indeed, producing informed and active citizens has long been an expressed mandate of the U.S. educational system. Still, much social science research, and recent rhetoric on college benefits, focuses on economic rather than civic returns to college. We find that college attendance increases self-reported voting and that most of that effect does not operate through degree completion or college-induced increases in socioeconomic status. In other words, civic returns to college do not hinge on its socioeconomic returns; instead, they appear to stem primarily from the college experience itself. Moreover, college attendance has a larger direct effect on voting for individuals who are less likely to attend college, i.e., those from relatively disadvantaged backgrounds. Our findings thus highlight that an exclusive focus on the economic benefits of college can mask the broader societal benefits of expanding higher education to disadvantaged youth.

REFERENCES

Ansolabehere, Stephen and Eitan Hersh. 2012. "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis* 20:437–59.

Amato, Paul R. 2010. "Research on Divorce: Continuing Trends and New Developments." *Journal of Marriage and Family* 72 (3): 650–66.

Arum, Richard, and Josipa Roksa. 2010. *Academically Adrift: Limited Learning on College*
31

Campuses. Chicago: University of Chicago Press.

Avin, Chen, Ilya Shpitser, and Judea Pearl. 2005. "Identifiability of Path-Specific Effects." *In Proceedings of International Joint Conference on Artificial Intelligence*, 357–63.

Baron, Reuben M, and David A Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51 (6): 1173–82.

Bartels, Larry M. 2018. *Unequal Democracy: The Political Economy of the New Gilded Age.* Princeton: Princeton University Press.

Berent, Matthew K., Jon A. Krosnick, and Arthur Lupia. 2016. "Measuring Voter Registration and Turnout in Surveys: Do Official Government Records Yield More Accurate Assessments?" *Public Opinion Quarterly* 80(3):597–621.

Berinsky, Adam J, and Gabriel S. Lenz. 2010. "Education and Political Participation: Exploring the Causal Link." *Political Behavior* 33: 357–73.

Bowman, Nicholas A. 2011. "Promoting Participation in a Diverse Democracy: A Meta-Analysis of College Diversity Experiences and Civic Engagement." *Review of Educational Research* 81 (1): 29–68.

Brady, Henry E, Sidney Verba, and Kay Lehman Schlozman. 1995. "Beyond SES: A Resource Model of Political Participation." *The American Political Science Review* 89 (2): 271–94.

Brand, Jennie E. 2010. "Civic Returns to Higher Education: A Note on Heterogeneous Effects." *Social Forces* 89 (2): 417–33.

Brand, Jennie. [forthcoming]. "Heterogeneous Effects of College on Reducing Poverty." in *The Inequality Reader, 5th edition*, David Grusky, Nima Dahir, and Claire Daviss eds., Routledge Press.

Brand, Jennie E. and Dwight Davis. 2011. "The Impact of College Education on Fertility: Evidence for Heterogeneous Effects." *Demography* 48(3):863-887.

Brand, Jennie E, Ravaris Moore, Xi Song, and Yu Xie. 2019. "Parental Divorce Is Not Uniformly Disruptive to Children's Educational Attainment." *Proceedings of the National Academy of Sciences* 116 (15): 7266–71.

Brand, Jennie E, and Yu Xie. 2010. "Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education." *American*

32

Sociological Review 75 (2): 273–302.

Brand, Jennie E., Jiahui Xu, Bernard Koch, and Pablo Geraldo. 2021. "Uncovering Sociological Effect Heterogeneity Using Tree-Based Machine Learning." *Sociological Methodology* 51(2):189–223.

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (5): 1127–60.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil. 2011. "Estimating Marginal Returns to Education." *The American Economic Review* 101 (6): 2754–81.

Cassel, Carol A, and Celia C Lo. 1997. "Theories of Political Literacy." *Political Behavior* 19 (4): 317–35.

Chatfield, Sara, and John Henderson. 2016. "Untangling the Education Effect: Moving Educational Interventions into the Experimental Frontier." In *New Advances in the Study of Civic Voluntarism: Resources, Engagement, and Recruitment*, edited by Casey Klofstad, 225–49. Philadelphia, PA: Temple University Press.

Cheng, Siwei, Jennie E. Brand, Xiang Zhou, Yu Xie, and Michael Hout. 2021. "Heterogeneous Returns to College over the Life Cycle." *Science Advances* 7, ebg7641.

Debell, Matthew, Jon A. Krosnick, Katie Gera, David S. Yeager, and Michael P. Mcdonald. 2020. "The Turnout Gap in Surveys: Explanations and Solutions." *Sociological Methods & Research* 49(4):1133–62.

Dee, Thomas S. 2004. "Are There Civic Returns to Education?" *Journal of Public Economics* 88 (9–10): 1697–1720.

Doyle, William R. and Benjamin T. Skinner. 2017. "Does Postsecondary Education Result in Civic Benefits?" *The Journal of Higher Education* 88(6):863–93.

Enamorado, Ted and Kosuke Imai. 2019. "Validating Self-Reported Turnout By Linking Public Opinion Surveys with Administration Records." *Public Opinion Quarterly* 83(4):723–48.

Freeman, Richard B. 2003. "What, Me Vote?" *NBER Working Paper*, no. 9896.

Friedman, Jerome. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *The Annals of Statistics* 29 (5): 1189–1232.

33

Galston, William A. 2001. "Political Knowledge, Political Engagement, and Civic Education." *Annual Review of Political Science* 4: 217–34.

Glynn, Carroll J., Michael E. Huge, and Carole A. Lunney. 2009. "The Influence of Perceived Social Norms on College Students' Intention to Vote." *Political Communication* 26 (1): 48–64.

Goldberg, Andreas C. and Pascal Sciarini. 2019. "Who Gets Lost, and What Difference Does It Make? Mixed Modes, Nonresponse Follow-up Surveys and the Estimation of Turnout." *Journal of Survey Statistics and Methodology* 7(4):520–44.

Hansen, Eric R, and Andrew Tyner. 2019. "Educational Attainment and Social Norms of Voting." *Political Behavior*.

Hastie, T.J., 2017. "Generalized Additive Models. " Pp. 249–307 in *Statistical Models in S*, edited by J. M. Chambers and T. J. Hastie. Routledge.

Hauser, Seth M. 2000. "Education, Ability, and Civic Engagement in the Contemporary United States." *Social Science Research* 29 (4): 556–82.

Heckman, James J., John Eric Humphries, and Gregory Veramendi. 2018. "The Nonmarket Benefits of Education and Ability." *Journal of Human Capital* 12 (2): 282–304.

Highton, Benjamin. 2009. "Revisiting the Relationship Between Educational Attainment and Political Sophistication." *The Journal of Politics* 71 (4): 1564–76.

Hillygus, D. Sunshine. 2005. "The Missing Link: Exploring the Relationship between Higher Education and Political Engagement." *Political Behavior* 27 (1): 25–47.

Hout, Michael. 2012. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology* 38: 379–400.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010. "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science* 25 (1): 51–71.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press.

Jack, Anthony A. 2019. The Privileged Poor: How Elite Colleges Are Failing Disadvantaged Students. Cambridge, MA: Harvard University Press.

Kam, Cindy D, and Carl L Palmer. 2008. "Reconsidering the Effects of Education on Political

Participation." *The Journal of Politics* 70 (3): 612–31.

Killewald, Alexandra, and Margaret Gough. 2013. "Does Specialization Explain Marriage Penalties and Premiums?" *American Sociological Review* 78 (3): 477–502.

Kingston, Paul W., Ryan Hubbard, Brent Lapp, Paul Schroeder, and Julia Wilson. 2003. "Why Education Matters." *Sociology of Education* 76 (1): 53–70.

Labaree, David F. 1997. "Public Goods Private Goods." *American Educational Research Journal* 34(1): 39-81.

Lahtinen, Hannu, Pekka Martikainen, Mikko Mattila, Hanna Wass, and Lauri Rapeli. 2019. "Do Surveys Overestimate or Underestimate Socioeconomic Differences in Voter Turnout? Evidence from Administrative Registers." *Public Opinion Quarterly* 83(2):363–85.

Leighley, Jan E., and Jonathan Nagler. 2013. *Who Votes Now?: Demographics, Issues, Inequality, and Turnout in the United States*. Princeton: Princeton University Press.

Lundberg, Shelly, Robert A Pollak, and Jenna Stearns. 2016. "Family Inequality: Diverging Patterns in Marriage, Cohabitation, and Childbearing." *Journal of Economic Perspectives* 30: 79–102.

Mayer, Alexander K. 2011. "Does Education Increase Political Participation?" *The Journal of Politics* 73 (03): 633–45.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos. 2004. "Does Education Improve Citizenship? Evidence from the U.S. and the U.K." *NBER Working Paper*, no. 9584.

Musick, Kelly, Jennie E. Brand, and Dwight Davis. 2012. "Variation in the Relationship between Education and Marriage: Marriage Market Mismatch?" *Journal of Marriage and Family* 74 (1): 53–69.

Olsen, Marvin E. 1972. "Social Participation and Voting Turnout: A Multivariate Analysis." *American Sociological Review* 37 (3): 317–33.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference Second Edition*. 2nd ed. Cambridge: Cambridge University Press.

Pew Research Center. 2018. *Commercial Voter Files and the Study of U.S. Politics*. https://www.pewresearch.org/methods/2018/02/15/commercial-voter-files-and-the-study-of-u-s-politics/.

Phinney, Robin. 2013. "Exploring Residential Mobility among Low-Income Families." *Social*

Service Review 87 (4): 780–815.

Putnam, Robert D. 2001. *Bowling Alone*. *Bowling Alone*. New York, NY: Simon and Schuster.

Quintelier, Ellen. 2010. "The Effect of Schools on Political Participation: A Multilevel Logistic Analysis." *Research Papers in Education* 25 (2): 137–54.

Ridgeway, Greg. 2019. "Generalized Boosted Models: A Guide to the Gbm Package." 2019. https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf.

Rolfe, Meredith. 2012. *Voter Turnout: A Social Theory of Political Participation*. Cambridge, UK: Cambridge University Press.

Schneider, Daniel. 2011. "Wealth and the Marital Divide." *American Journal of Sociology* 117 (2): 627–67.

Schudde, Lauren T. 2011. "The Causal Effect of Campus Residency on College Student Retention." *The Review of Higher Education* 34(4):581–610.

Selb, Peter and Simon Munzert. 2013. "Voter Overrepresentation, Vote Misreporting, and Turnout Bias in Postelection Surveys." *Electoral Studies* 32(1):186–96.

Sewell, William H, and Robert M. Hauser. 1975. *Education, Occupation, and Earnings*. New York: Academic Press.

Skinner, Benjamin T. and William R. Doyle. 2021. "Do Civic Returns to Higher Education Differ across Subpopulations? An Analysis Using Propensity Forests." *Journal of Education Finance* 46(4):519–62.

Smock, Pamela J, Wendy D Manning, and Sanjiv Gupta. 1999. "The Effect of Marriage and Divorce on Women's Economic Well-Being." *American Sociological Review* 64 (6): 794–812.

Stevens, Mitchell L., Elizabeth A. Armstrong, and Richard Arum. 2008. "Sieve, Incubator, Temple, Hub: Empirical and Theoretical Advances in the Sociology of Higher Education." *Annual Review of Sociology* 34 (1): 127–51.

Tenn, Steven. 2007. "The Effect of Education on Voter Turnout." *Political Analysis* 15 (4): 446–64.

U.S. Bureau of the Census. 2012. "Table 10. Reasons For Not Voting, By Selected Characteristics: November 2012."

———. 2018a. "Table 7. Reported Voting and Registration of Family Members, by Age and Family Income: November 2018."

———. 2018b. "Table A-1. Reported Voting and Registration by Race, Hispanic Origin, Sex, and Age Groups: November 1964 to 2018."

———. 2018c. "Table A-2. Reported Voting and Registration by Region, Educational Attainment and Labor Force Status for the Population 18 and Over: November 1964 to 2018."

VanderWeele, Tyler J. 2010. "Bias Formulas for Sensitivity Analysis for Direct and Indirect Effects NIH Public Access." *Epidemiology* 21 (4): 540–51.

———. 2015. *Explanation in Causal Inference*. Oxford: Oxford University Press.

VanderWeele, Tyler J, and Onyebuchi A Arah. 2011. "Unmeasured Confounding for General Outcomes, Treatments, and Confounders: Bias Formulas for Sensitivity Analysis." *Epidemiology* 22 (1): 42–52.

Vanderweele, Tyler J, Linda Valeri, and Elizabeth L Ogburn. 2012. "The Role of Measurement Error and Misclassification in Mediation Analysis Mediation and Non-Differential Measurement Error NIH Public Access." *Epidemiology* 23 (4): 561–64.

Wolak, Jennifer. 2018. "Feelings of Political Efficacy in the Fifty States." *Political Behavior* 40:763–84.

Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes*. New Haven: Yale University Press.

Zhou, Xiang. 2022. "Semiparametric Estimation for Causal Mediation Analysis with Multiple Causally Ordered Mediators" *Journal of the Royal Statistical Society: Series B.*

Zhou, Xiang, and Yu Xie. 2020. "Heterogeneous Treatment Effects in the Presence of Self-Selection: A Propensity Score Perspective." *Sociological Methodology* 50 (1):350-385

Zhou, Xiang, and Teppei Yamamoto. 2022. "Tracing Causal Paths from Experimental and Observational Data." *The Journal of Politics*.

TABLES AND FIGURES

Table 1: Descriptive Statistics of Outcome, Covariates, and Mediators by College Attendance, NLSY79

	Full Analytic Sample		Non-College Attenders		College Attenders	
	Mean	SD	Mean	SD	Mean	SD
Outcome						
Voted in 2006 election (binary 0/1)	0.69	-	0.61	-	0.84	-
Pre-College Covariates						
Male (binary 0/1)	0.50	-	0.50	-	0.48	-
Black (binary 0/1)	0.14	-	0.16	-	0.11	-
Hispanic (binary 0/1)	0.05	-	0.06	-	0.04	-
Father's Education (0-20)	12.24	3.26	11.33	2.94	13.73	3.21
Mother's Education (0-20)	11.98	2.40	11.34	2.15	13.04	2.41
Don't Know Father's Education (binary 0/1)	0.04	-	0.06	-	0.02	-
Father White Collar Occupation (binary 0/1)	0.27	-	0.17	-	0.44	-
Parental Income 1979 (0-750)	220.22	127.63	192.68	110.07	265.22	140.87
Lived with Both Parents at 14 (binary 0/1)	0.76	-	0.73	-	0.82	-
Number of Siblings (0-15)	3.02	2.06	3.24	2.20	2.66	1.76
Catholic Background (binary 0/1)	0.32	-	0.32	-	0.34	-
Jewish Background (binary 0/1)	0.01	-	0.01	-	0.02	-
Southern Residence (binary 0/1)	0.32	-	0.32	-	0.32	-
Rural Residence (binary 0/1)	0.23	-	0.25	-	0.20	-
ASVAB Ability Score (-2.5-2.4)	0.13	0.68	-0.11	0.62	0.53	0.56
High School College Prep Program (binary 0/1)	0.35	-	0.21	-	0.59	-
College Aspirations (binary 0/1)	0.58	-	0.41	-	0.85	-
College Expectations (binary 0/1)	0.46	-	0.27	-	0.78	-

Closest Friend Aspires to College (binary 0/1)	0.48	-	0.34	-	0.70	
Rotter Locus of Control (4-16)	8.78	2.27	9.05	2.22	8.34	2.28
Rosenberg Self-Esteem (8-30)	22.22	3.84	21.74	3.70	22.95	3.95
High on Delinquency Scale (binary 0/1)	0.79	-	0.82	-	0.74	-
School Disadvantage (0-99)	18.51	16.13	21.02	16.91	14.43	13.82
Percent Students Black/Hispanic (0-100)	19.45	22.62	21.46	23.79	16.17	20.15
High on Scale of Traditional Family Values (binary 0/1)	0.21	-	0.23	-	0.16	-
Married by Age 18 (binary 0/1)	0.02	-	0.03	-	0.00	-
Parent by Age 18 (binary 0/1)	0.02	-	0.02	-	0.00	-
Family Formation Mediators						
Prop of Time Married 1990-1998 (0-1)	0.58	0.41	0.56	0.42	0.61	0.41
Prop of Time Unmarried Parent 1990-1998 (0-1)	0.16	0.31	0.23	0.35	0.06	0.19
Number of Family Transitions 1990-1998 (0-14)	0.80	1.72	1.06	1.95	0.37	1.15
Socioeconomic Status Mediators						
Log of Mean Income, Age 35-40 (2.0-12.0)	10.19	0.89	9.98	0.87	10.51	0.85
Prop of Time Out of Labor Force 1998-2006 (0-0.2)	0.03	0.04	0.03	0.04	0.02	0.04
Prop of Time in Poverty 1998-2006 (binary 0/1)	0.08	0.21	0.11	0.24	0.03	0.10
Cum Welfare Received 1998-2006 (2014 $) (0-150000)	2170	10955	3141	13266	586.88	4968
Count of Jobs Held 1990-2006 (1-40)	4.44	3.36	4.61	3.44	4.16	3.22
Job Tenure at Age 40 (0-1446)	322	314	311	317	340	310
Count of Moved Addresses 2000-2006 (0-16)	1.12	1.61	1.23	1.70	0.96	1.43
No Home Ownership 2004 (binary 0/1)	0.26	-	0.32	-	0.17	-
N	2961		1950		1011	

Notes: All descriptive statistics are weighted with the NLSY79 sample weight. Unless noted as binary, all variables are continuous.

Dashes are used in place of standard deviations for binary variables.

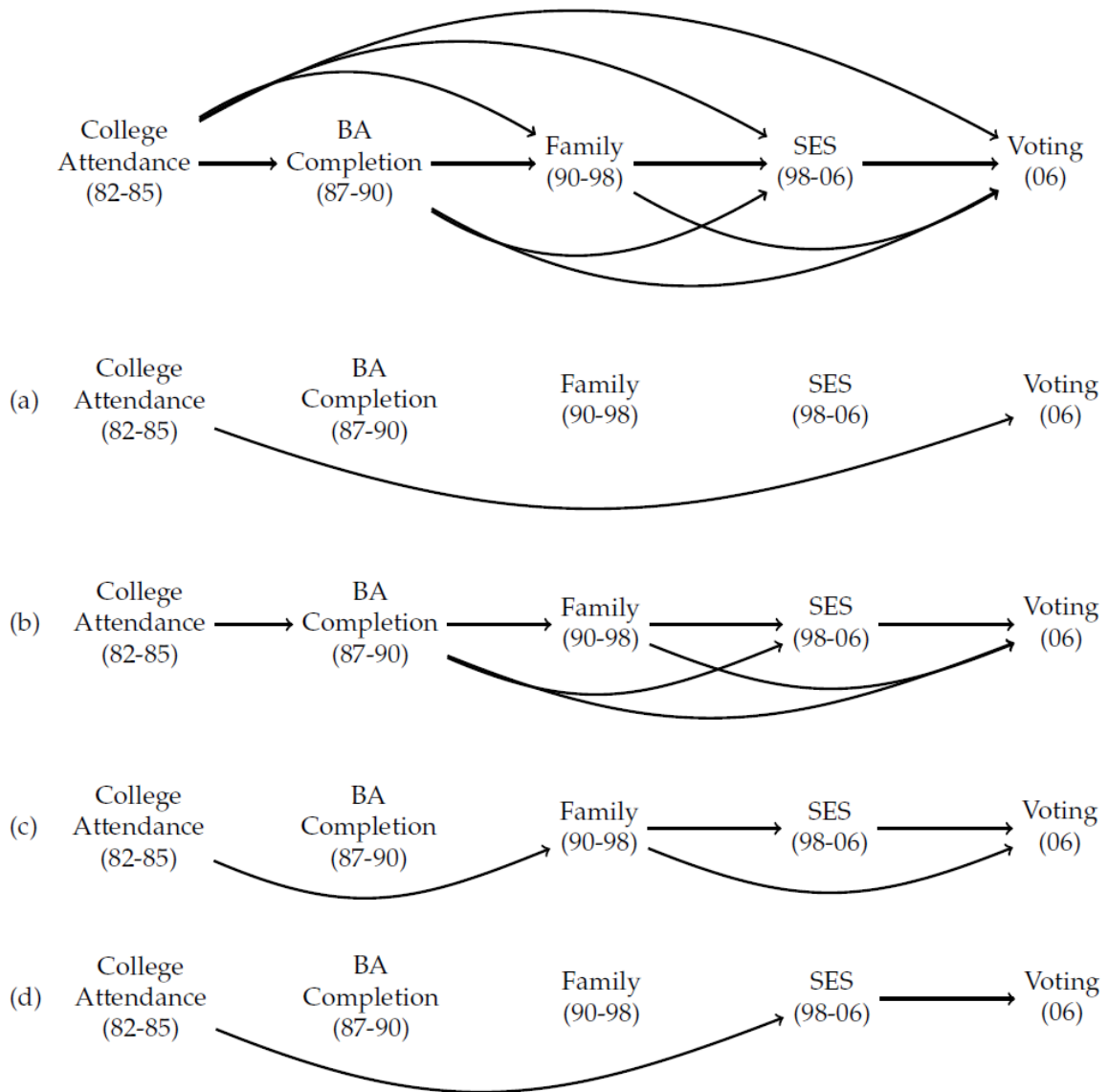[a]Rotter locus of control is scaled so that higher values indicate less control.

Figure 1: Causal Pathways Through Which College May Affect Voting.
Notes: Numbers in parentheses indicate the years in which the corresponding variables are measured.
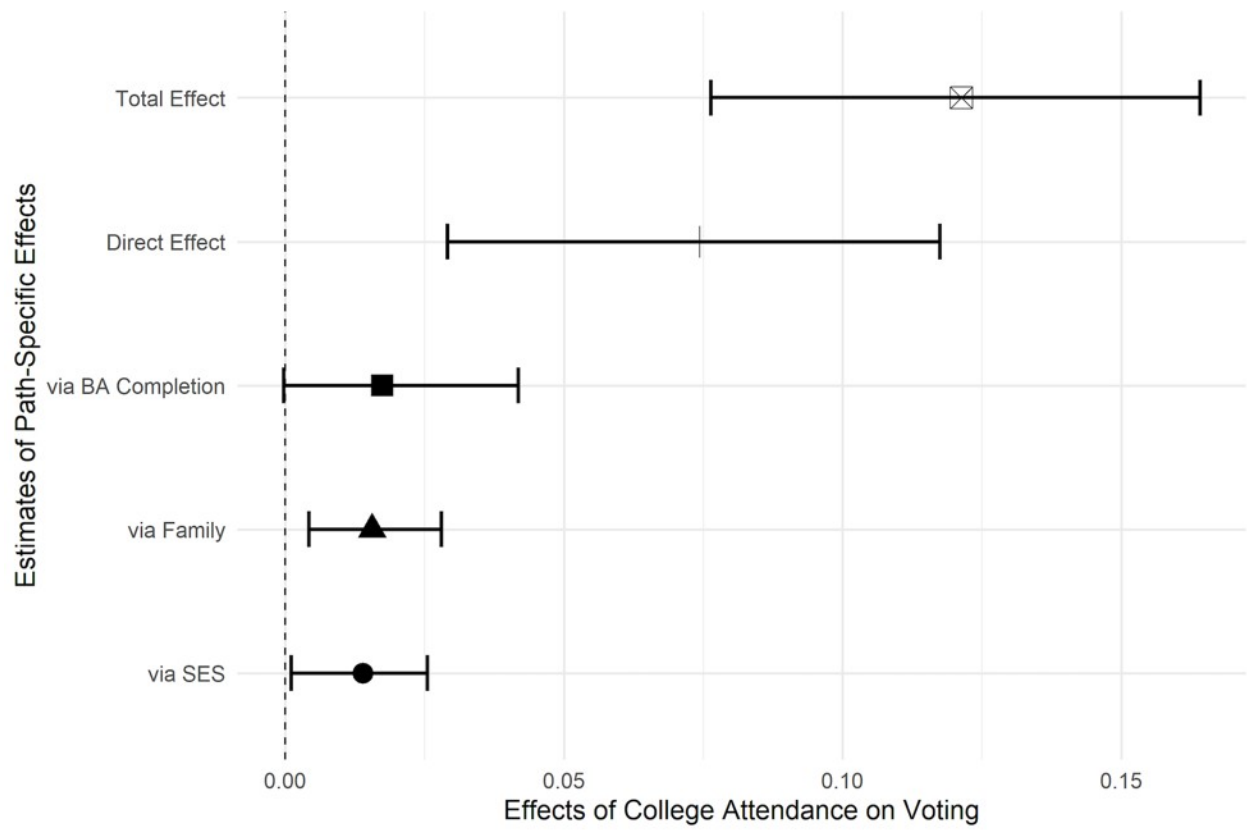
Figure 2: Estimated Path-Specific Effects of College Attendance on Voting With 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY79
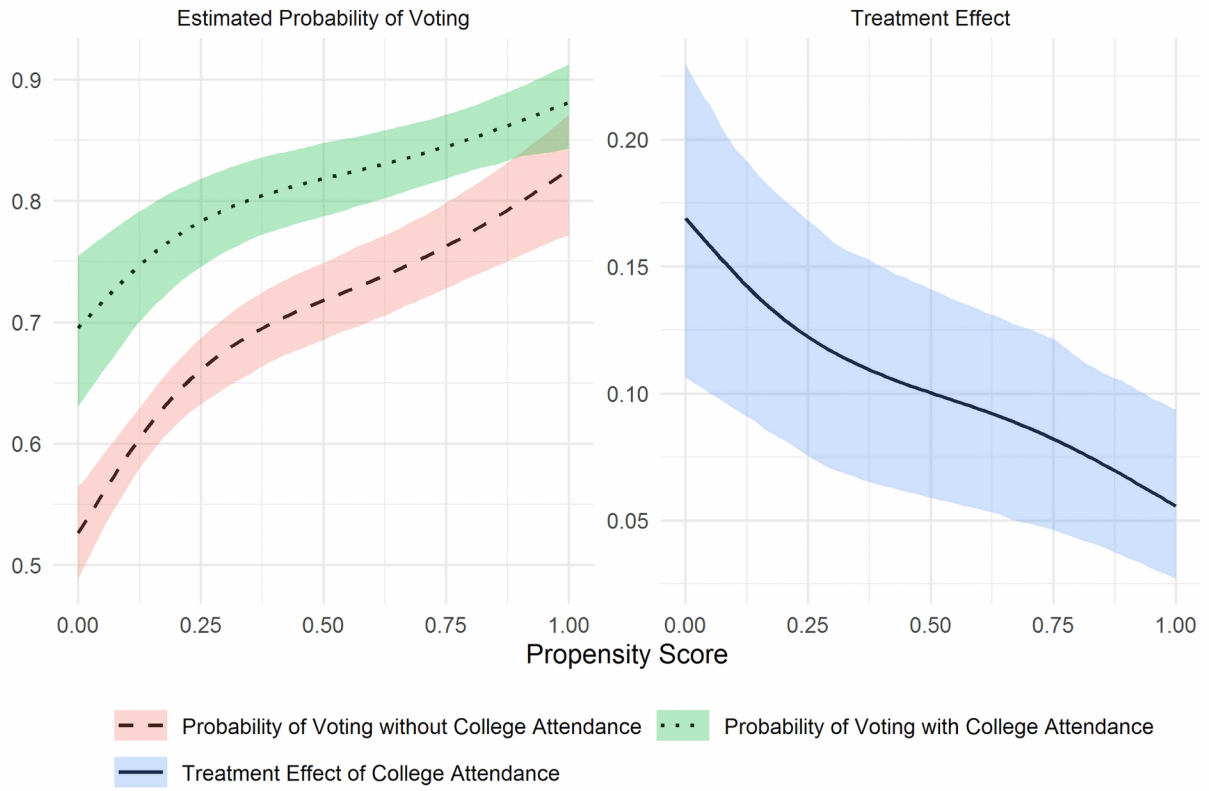
Figure 3: Heterogeneous Total Effects of College Attendance on Voting by Propensity Scores with 95% Bootstrap Confidence Bands (1,000 iterations), NLSY79
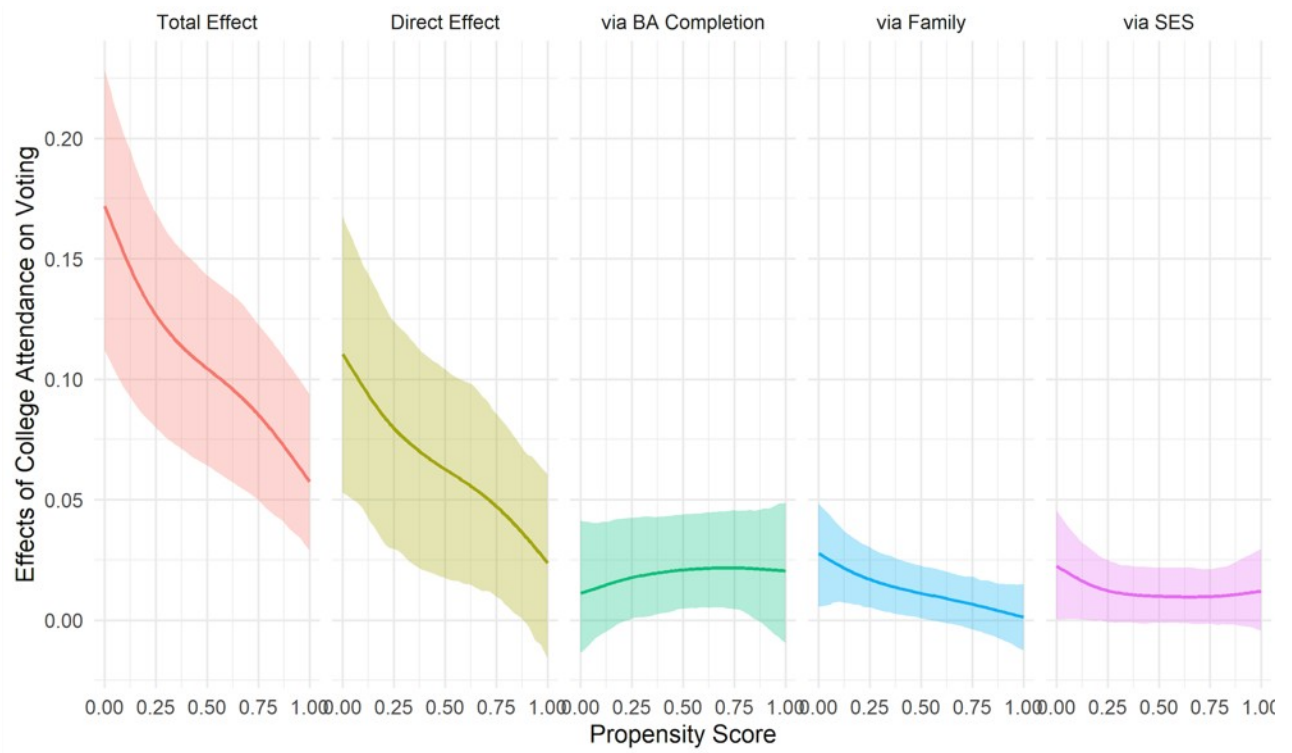
Figure 4: Heterogeneous Total, Direct, and Indirect Effects of College Attendance on Voting by Propensity Scores with 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY79
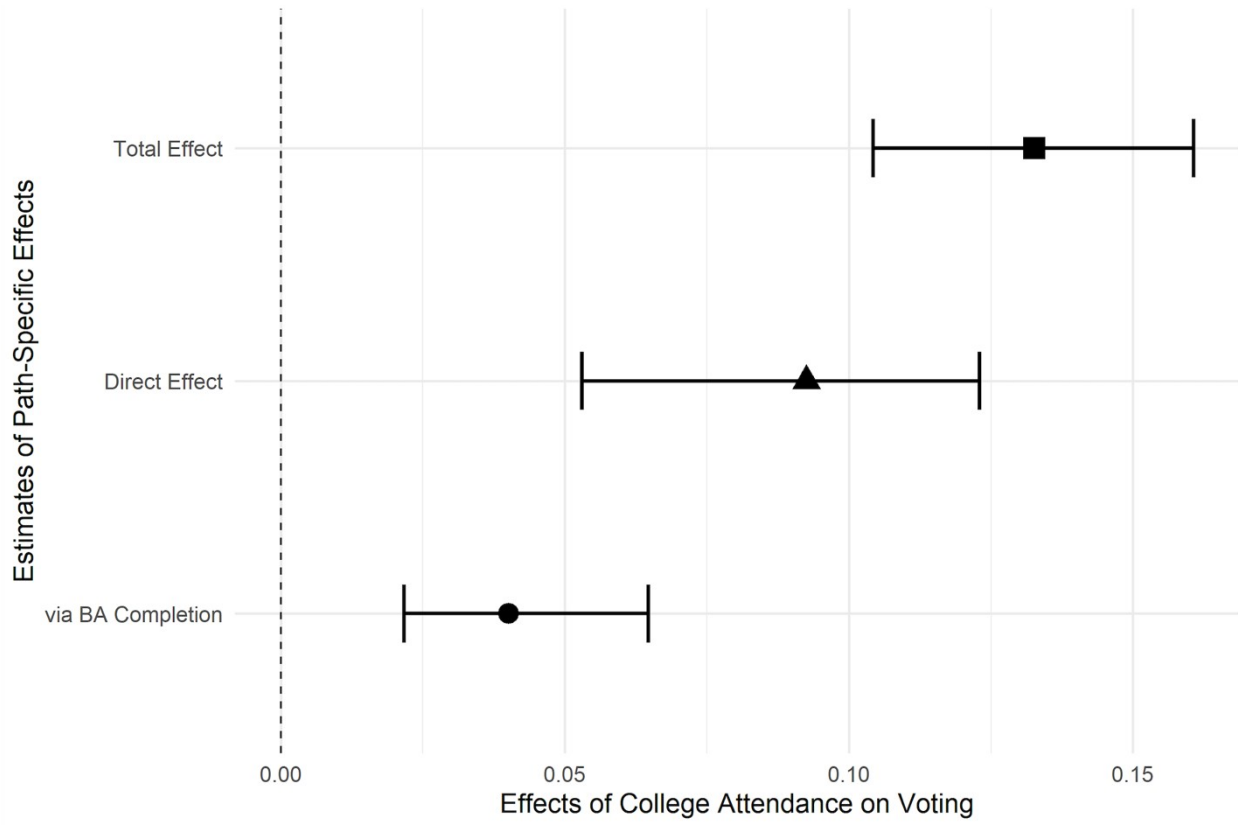
Figure 5: Estimated Path-Specific Effects of College Attendance on Voting With 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY97
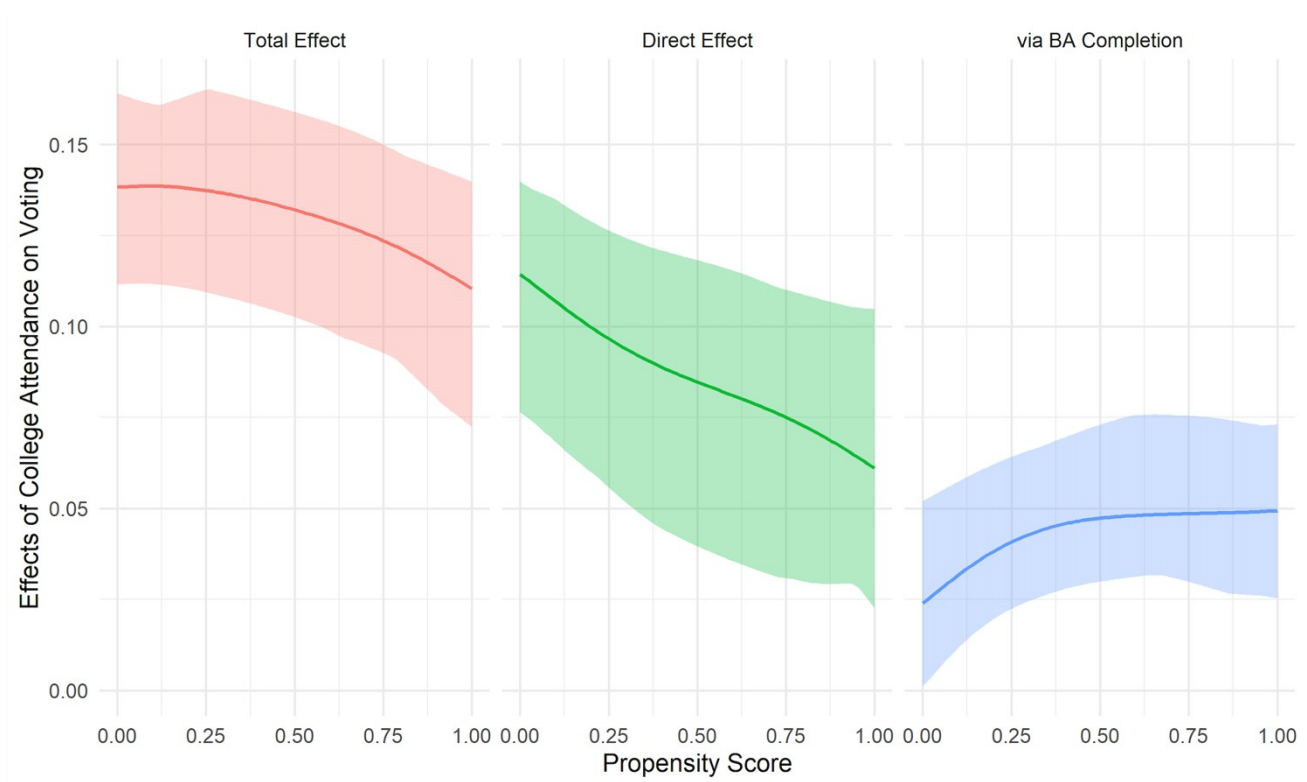
Figure 6: Heterogeneous Total, Direct, and Indirect Effects of College Attendance on Voting by Propensity Scores with 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY97

Table A.1: Descriptive Statistics of Outcome, Covariates, and Mediators by College Attendance, NSLY97

	Full Analytic Sample		Non-College Attenders		College Attenders	
	Mean	SD	Mean	SD	Mean	SD
Outcome						
Voted in 2010 election (binary 0/1)	0.40	-	0.31	-	0.54	-
Pre-College Covariates						
Male (binary 0/1)	0.51	-	0.54	-	0.45	-
Black (binary 0/1)	0.16	-	0.19	-	0.12	-
Hispanic (binary 0/1)	0.13	-	0.17	-	0.08	-
Father's Education (0-20)	12.81	3.03	11.94	2.71	14.28	2.98
Mother's Education (0-20)	12.83	2.74	12.07	2.49	14.11	2.66
Parental Income 1997 (0-246000)	50597	40387	39764	31251	68768	46925
Lived with Both Parents at 14 (binary 0/1)	0.52	-	0.42	-	0.69	-
Number of Siblings (0-45)	2.54	2.13	2.81	2.31	2.09	1.7
Catholic Background (binary 0/1)	0.26	-	0.25	-	0.27	-
Jewish Background (binary 0/1)	0.01	-	0.01	-	0.03	-
Southern Residence (binary 0/1)	0.34	-	0.35	-	0.31	-
Rural Residence (binary 0/1)	0.21	-	0.20	-	0.21	-
ASVAB Ability Score (-3.2-2.6)	-0.20	0.65	-0.40	0.60	0.14	0.60
Self-Reported High School GPA (0-4)	2.84	0.82	2.54	0.79	3.34	0.60
High School College Prep Program (binary 0/1)	0.32	-	0.18	-	0.56	-
Believes School is Safe	0.33	-	0.27	-	0.44	-
Believes Teachers are Good (binary 0/1)	0.18	-	0.15	-	0.24	-
Friend's College Completion Aspirations (binary 0/1)	0.22	-	0.18	-	0.30	-
High on Delinquency Scale (binary 0/1)	0.55	-	0.62	-	0.44	-

Married by Age 18 (binary 0/1)	0.02	-	0.03	-	0.00	-
Parent by Age 18 (binary 0/1)	0.07	-	0.10	-	0.01	-
N	7620		5073		2547	

Notes: All descriptive statistics are weighted with the NLSY97 sample weight. Unless noted as binary, all variables are continuous. Dashes are used in place of standard deviations for binary variables.

APPENDIX B: IDENTIFICATION OF PATH-SPECIFIC EFFECTS

Let A denote treatment (i.e., four-year college attendance), Y the outcome of interest (i.e., voted in 2006), X the vector of pretreatment covariates, C the four-year college completion mediator, L the set of mediators reflecting family formation and stability, and M the set of mediators reflecting SES. The causal paths (a), (b), (c), (d) shown in Figure 1 can be represented as (a) $A \rightarrow Y$; (b) $A \rightarrow CY$; (c) $A \rightarrow LY$; and (d) $A \rightarrow M \rightarrow Y$. As shown in Zhou and Yamamoto (2022), under the assumption of no unobserved confounding, the corresponding path-specific effects can be identified as

$$\tau_{A \rightarrow Y} = E\big[E\big[E[Y|X, A=1, C, L, M]\big|X, A=0\big] - E[Y|X, A=0]\big], (B.1)$$

$$\tau_{A \rightarrow M \rightarrow Y} = E\big[E\big[E[Y|X, A=1, C, L]\big|X, A=0\big] - E\big[E[Y|X, A=1, C, L, M]\big|X, A=0\big]\big], (B.2)$$

$$\tau_{A \rightarrow LY} = E\big[E\big[E[Y|X, A=1, C]\big|X, A=0\big] - E\big[E[Y|X, A=1, C, L]\big|X, A=0\big]\big], (B.3)$$

$$\tau_{A \rightarrow CY} = E\big[E[Y|X, A=1] - E\big[E[Y|X, A=1, C]\big|X, A=0\big]\big]. (B.4)$$

To gain some intuition of how the above formulas work, let us consider equation (B.1). Here, the inner expectation $E[X, A=1, C, L, M]$ captures the predicted voting outcome given pretreatment covariates X, college completion status C, family characteristics L, and SES M had everyone, contrary to fact, attended college. The outer expectation $E[X, A=0]$ is the conditional mean of this predicted value given pretreatment covariates X among non-college goers. Thus, this term reflects a counterfactual of what might be expected for those who did not attend college had they done so, but with their mediators (college completion status, family characteristics, and SES) kept at their realized values under non-attendance ($A=0$). Then, by subtracting from this term the predicted voting outcome given pretreatment covariates X among non-college goers, we

48

obtain the direct effect of college given pretreatment covariates X, the mean of which (the outmost expectation) constitutes the average direct effect of college. Equations (B.2-B.4) can be understood analogously.
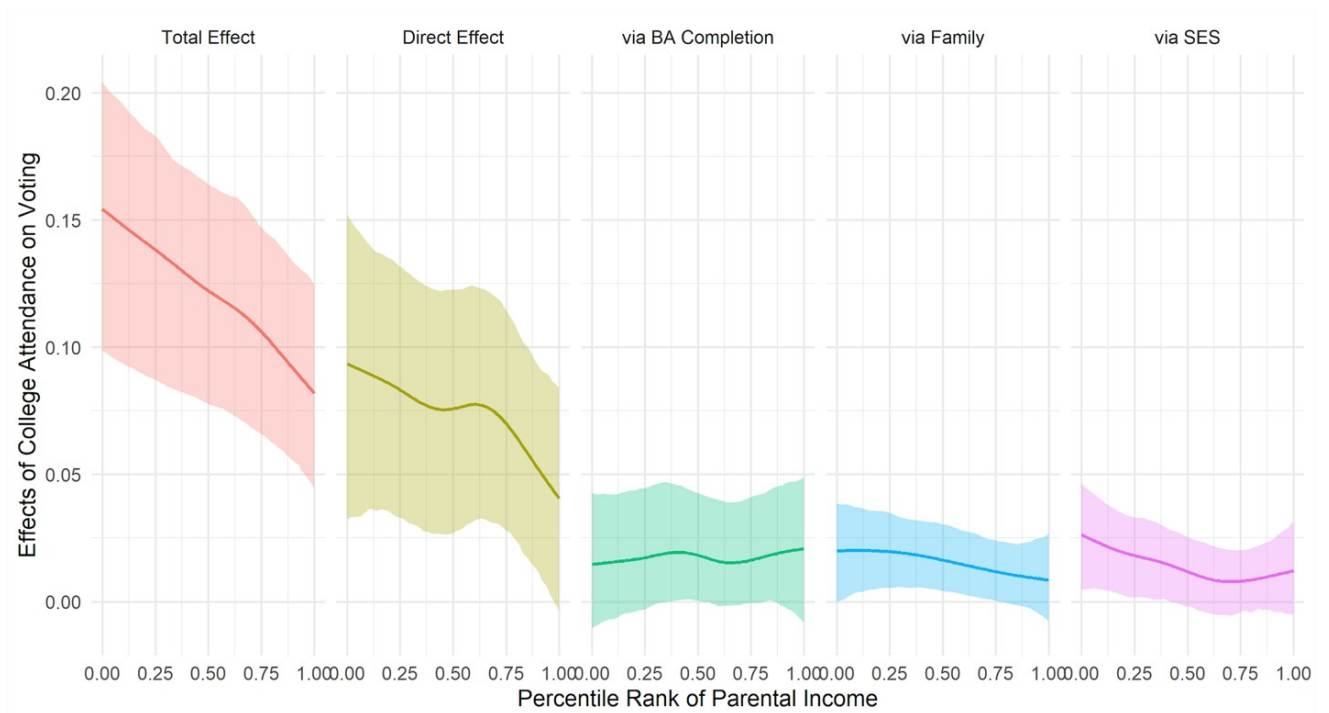
Figure C.1: Heterogeneous Total, Direct, and Indirect Effects of College Attendance on Voting by Parental Income with 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY79
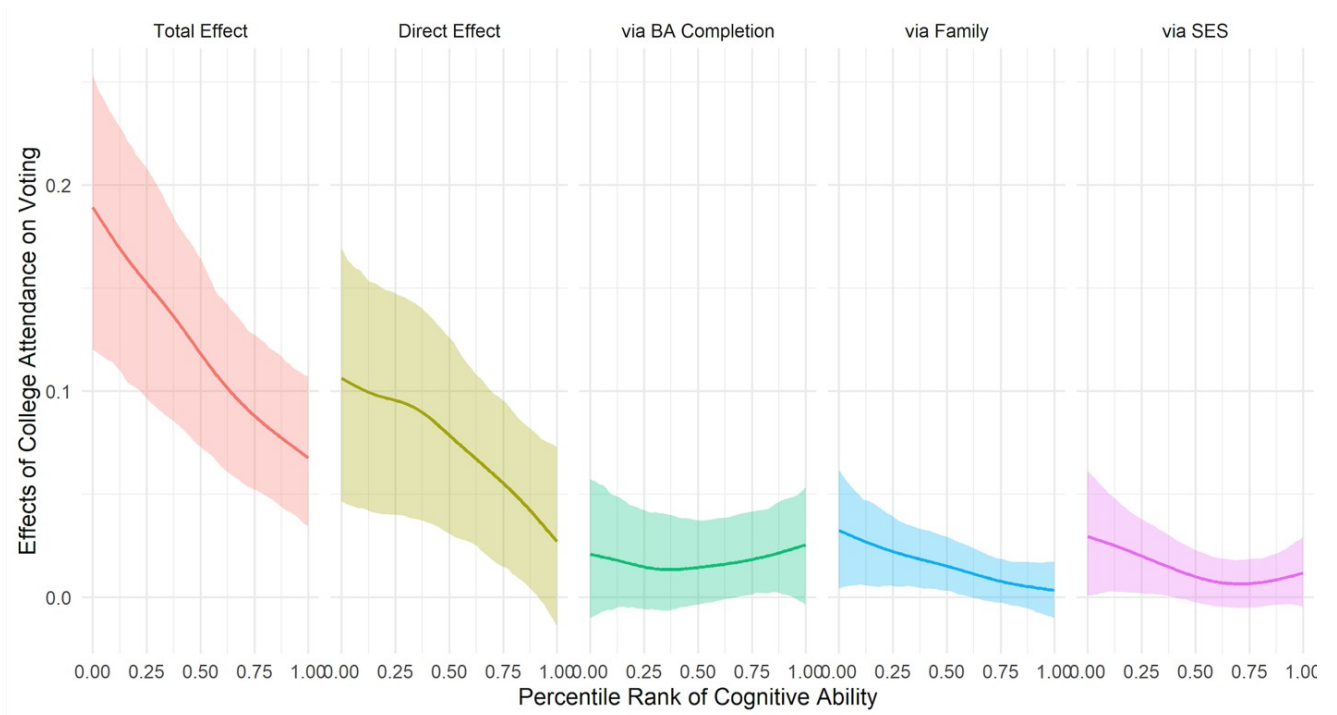
Figure C.2: Heterogeneous Total, Direct, and Indirect Effects of College Attendance on Voting by Cognitive Ability with 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY79

As we note in the main text, we replicated our main NLSY79 and NLSY97 analyses with an imputed measure of voting that addresses the potential impact of voter overrepresentation. Rather than omitting individuals who did not answer the voting item, we code all individuals who were missing data for the voting question (either due to missing in the wave or missing the specific item) as non-voters. This imputation assumption is grounded in previous research showing that individuals who do not participate in surveys tend to vote at lower rates than those who participate (Goldberg and Sciarini 2019; Lahtinen et al. 2019). Still, this is a very strong assumption and is likely to overrepresent non-voters.

We begin by replicating analyses for the NLSY79 sample. These results show some differences from the main results. First, the total effect of college on voting is about 8 percentage points (Figure D.1), compared to about 12 percentage points in the original sample. Second, the direct effect of college attendance on voting is smaller (about .03 compared to .08), and the mediating effect via college completion is larger (about .04 compared to .02) (Figure D.2). Finally, there remains a pattern of negative selection in the direct effect, but there is now a modest positive pattern in the mediating effects of college completion and SES. The corresponding results for the NLSY97 sample are shown in Appendix Figures D.3 and D.4. The overall estimated total and direct effects, and the mediating effect via college completion are all quite similar to the estimates from our main analyses, at around 13, 9, and 4 percentage points, respectively. However, the pattern of negative selection by the propensity score in the total effect is much less pronounced than in the non-missing sample (see Figure D.4). Yet, we still find a pattern of negative selection in the direct effect, and it appears that the lack of effect

heterogeneity in the total effect is due to a positive slope in the mediating effect via college completion.

      We also assess the potential bias due to unobserved confounding variables. First, we address potential bias in the treatment-outcome relationship. Although we have adjusted for an array of baseline covariates in our analyses, there may still be unobserved individual attributes that affect both college attendance and political participation. For simplicity, let us consider a binary unobserved confounder U, say some unmeasured personality trait. Under the simplifying assumption that the average effect of U on voting does not depend on treatment status A, the bias for the total effect conditional on the baseline covariates $X = x$ is given by

$$B_{total}(x) = \gamma \delta$$

where

$$\gamma = Pr[Y = 1 | U = 1, A, X = x] - Pr[Y = 1 | U = 0, A, X = x]$$

and

$$\delta = Pr[U = 1 | A = 1, X = x] - Pr[U = 1 | A = 0, X = x].$$

That is, γ is the difference in voting between individuals with and without unobserved characteristic U conditional on treatment status A and baseline covariates $X = x$, and δ is the difference in the prevalence of U between individuals with and without a college education conditional on baseline covariates $X = x$.

      Given that a characteristic that facilitates college attendance is likely positively associated with voting, consider the case where γ and δ are of the same sign. Here, the bias is positive, meaning the total effect of college attainment on voting is overestimated. Table D.1 shows the bias-adjusted estimates of the total effect across a range of plausible values for γ and δ. The estimated total effect of college attendance on voting is quite robust. Even if the

53

unobserved trait U increases voting by as much as 10 percentage points and the prevalence of U differs by as much as 20 percentage points between college and non-college goers, the bias-adjusted estimate of the total effect will still be statistically significant and substantively similar to the original estimate.

Second, unobserved confounders may exist for the mediator-outcome relationship. For example, neighborhood-level social capital, an unobserved variable in this study, may affect both the mediators and political participation. In this case, while the total effect estimate may still be unbiased, the direct and indirect effects will likely be overestimated or underestimated. To explore the direction and magnitude of potential bias, we again use a bias factor approach (VanderWeele 2010). Here, we consider a binary unobserved confounder U, say social capital, for the mediator-outcome relationships. Under some simplifying assumptions, the bias for the direct effect conditional on the baseline covariates $X=x$ is given by

$$B_{direct}(x)=\gamma\delta$$

where

$$\gamma=Pr[Y=1|U=1,X=x,A,C,L,M]-Pr[Y=1|U=0,X=x,A,C,L,M]$$

and

$$\delta=Pr[U=1|A=1,X=x,C,L,M]-Pr[U=1|A=0,X=x,C,L,M].$$

Here, γ is the difference in voting participation between individuals with and without social capital U conditional on college status A, the baseline covariates $X=x$, and the mediators reflecting college completion (C), family formation and stability (L), and socioeconomic attainment (M), while δ is the difference in the prevalence of U between individuals with and without college attendance conditional on baseline covariates and the mediators.

54

Considering that social capital generally boosts political participation, let us assume, without loss of generality, that $\gamma > 0$. The sign of δ, however, may be either positive or negative. Social capital tends to be higher among individuals who attended college. Yet if both college attendance and social capital promote college completion, family formation and stability, and socioeconomic status, conditioning on these mediators may reduce or even reverse their association. Thus, we assess the bias term $B_{direct}(x)$ for both positive and negative values of δ. Table D.2 shows the bias-adjusted estimates of the direct effect across a range of plausible values for γ and δ. The direct effect of college on voting will be underestimated if δ is negative and overestimated if δ is positive. Yet, our main conclusion that a substantial portion of the total effect of college attendance is direct is fairly robust. For example, if the unobserved confounder U increases voting by 10 percentage points and the prevalence of U differs by 10 percentage points between college and non-college goers with the same baseline characteristics and mediator values, the associated bias will only be 0.01, much smaller than the estimated direct effect.

The bias formulas presented above also inform the sensitivity of our results on treatment effect heterogeneity to unobserved selection. First, we note that if the γ and δ parameters are constant across different values of the propensity score, then low- and high-propensity individuals will suffer the same amount of bias. In this case, our finding that the effect of college on voting is larger among low-propensity individuals will be unchanged. However, the γ and δ parameters may differ between low- and high-propensity individuals. For example, the effect of an unmeasured characteristic on voting might be stronger among low-propensity individuals than among high-propensity individuals, for whom voting likely reflects a social norm and thus

depends less on personality traits. If so, our finding of negative selection might be a result of differential selection bias. If we use γ_L and δ_L to denote the γ and δ parameters defined above for low-propensity individuals and γ_H and δ_H for high-propensity individuals, the bias for the estimated difference in treatment effect between the two subpopulations will be

$$B_{diff} = \gamma_L \delta_L - \gamma_H \delta_H.$$

Our results indicated that the estimated total effect of college attendance decreases as the propensity score increases, ranging from .17 to .06. Thus, the difference in bias between low- and high-propensity individuals would have to be as large as .11 (.17-.06) to explain away effect heterogeneity. Given the range of plausible values for $\gamma_L \delta_L$ and $\gamma_H \delta_H$ (Table D.2), it is unlikely that unobserved confounding can explain away the degree of negative selection shown in Figure 3. The degree of negative selection in the estimated direct effect reported in Figure 4 likewise appears too large to be plausibly attributed to differential selection bias.
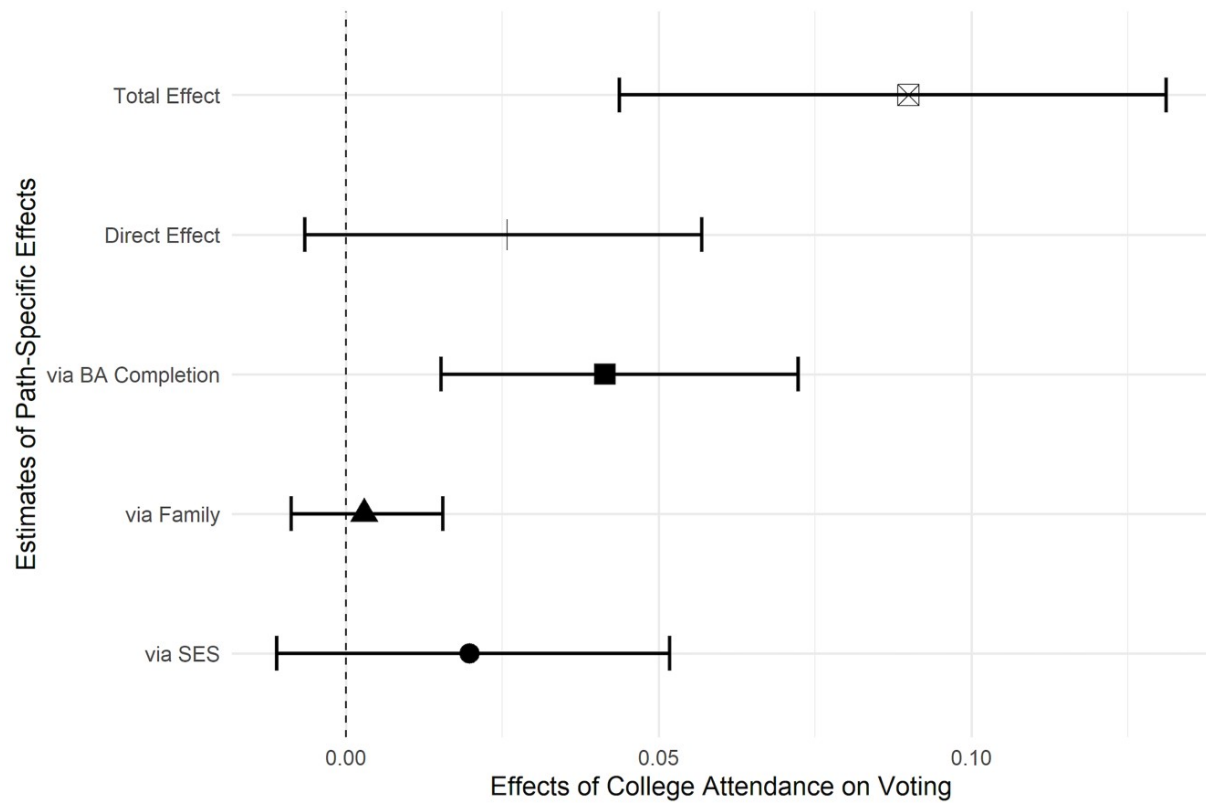
Figure D.1: Estimated Path-Specific Effects of College Attendance on Voting With 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY79 Overrepresentation Bias Sample
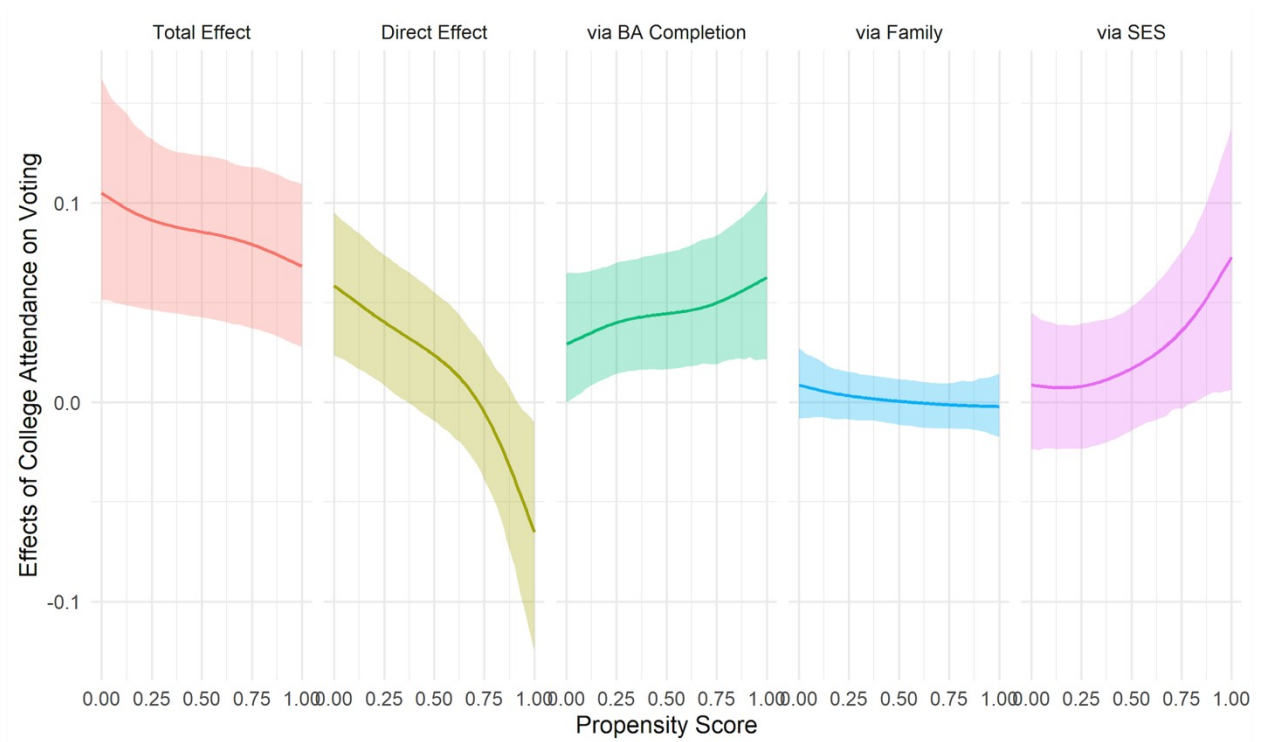
Figure D.2: Heterogeneous Total, Direct, and Indirect Effects of College Attendance on Voting by Propensity Scores with 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY79 NLSY79 Overrepresentation Bias Sample
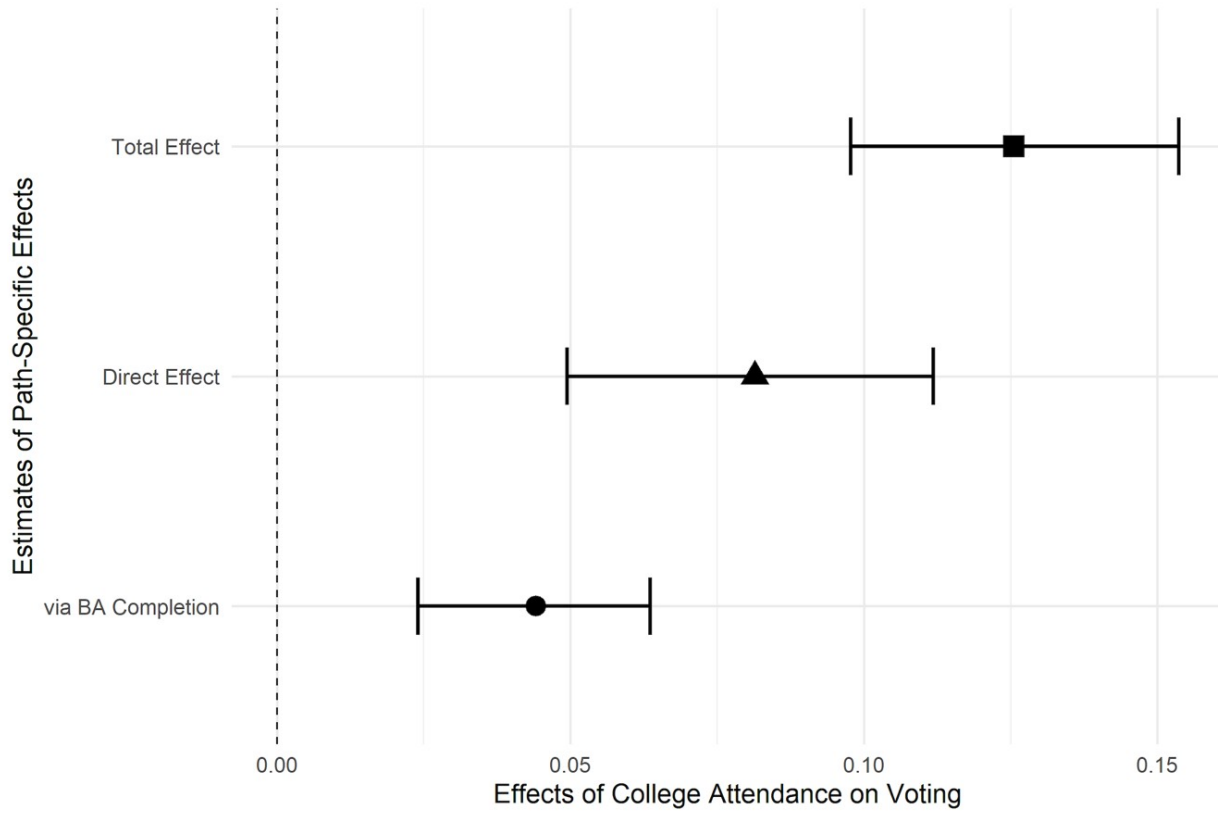
Figure D.3: Estimated Path-Specific Effects of College Attendance on Voting With 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY97 Overrepresentation Bias Sample
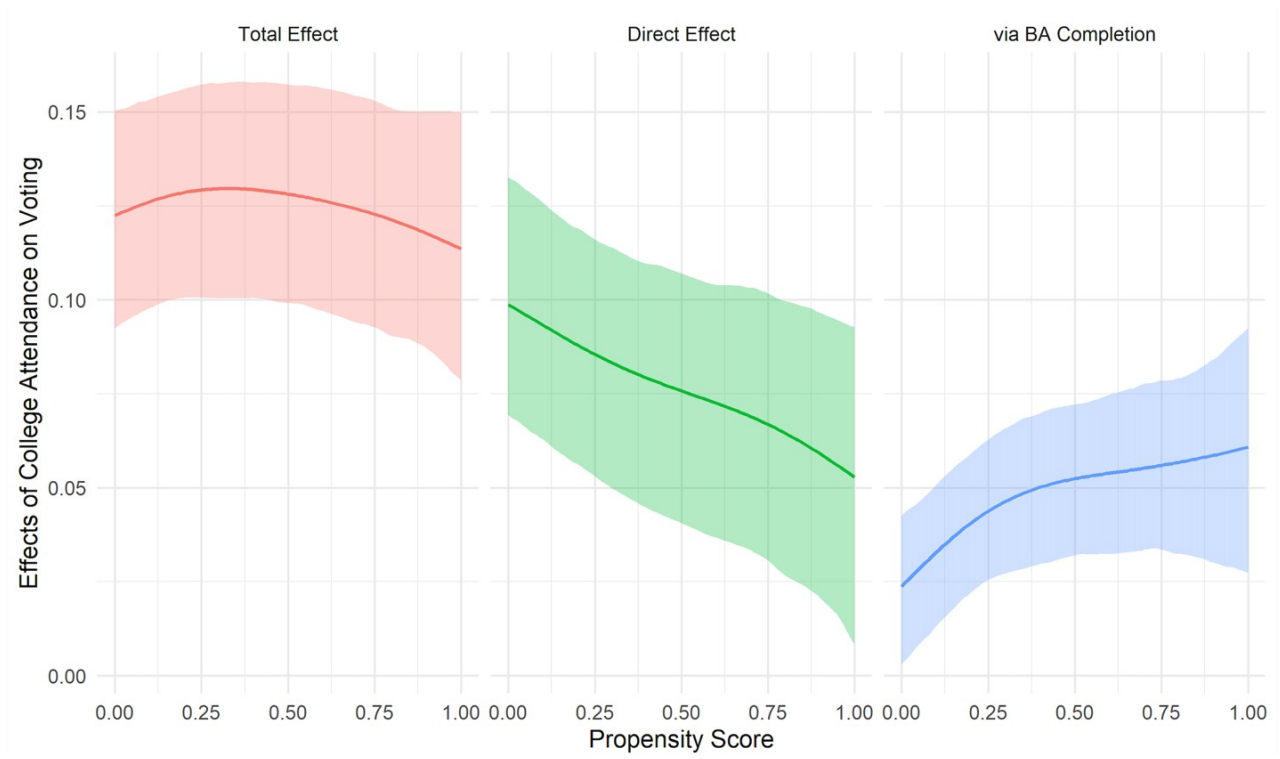
Figure D.4: Heterogeneous Total, Direct, and Indirect Effects of College Attendance on Voting by Propensity Scores with 95% Bootstrap Confidence Intervals (1,000 iterations), NLSY97 Overrepresentation Bias Sample

Table D.1: Sensitivity Results for Total Effects of College Attendance on Voting, NLSY79

Sensitivity Parameters		Bias	College Attendance	
γ	δ		Bias-adjusted Estimate	95% Confidence Interval
0.00	0.00	0.00	0.12	(0.08, 0.16)
0.05	0.05	0.00	0.12	(0.07, 0.16)
0.10	0.05	0.01	0.12	(0.07, 0.16)
0.20	0.05	0.01	0.11	(0.07, 0.15)
0.05	0.10	0.01	0.12	(0.07, 0.16)
0.10	0.10	0.01	0.11	(0.07, 0.15)
0.20	0.10	0.02	0.10	(0.06, 0.14)
0.05	0.20	0.01	0.11	(0.07, 0.15)
0.10	0.20	0.02	0.10	(0.06, 0.14)
0.20	0.20	0.04	0.08	(0.04, 0.12)

Table D.2: Sensitivity Results for Direct Effects of College Attendance on Voting, NLSY79

Sensitivity Parameters		Bias	College Attendance	
γ	δ		Bias-adjusted Estimate	95% Confidence Interval
0.00	0.00	0.00	0.07	(0.03, 0.12)
0.05	-0.05	0.00	0.08	(0.03, 0.12)
0.10	-0.05	-0.01	0.08	(0.03, 0.12)
0.20	-0.05	-0.01	0.08	(0.04, 0.13)
0.05	-0.10	-0.01	0.08	(0.03, 0.12)
0.10	-0.10	-0.01	0.08	(0.04, 0.13)
0.20	-0.10	-0.02	0.09	(0.05, 0.14)
0.05	0.05	0.00	0.07	(0.03, 0.11)
0.10	0.05	0.01	0.07	(0.02, 0.11)
0.20	0.05	0.01	0.06	(0.02, 0.11)
0.05	0.10	0.01	0.07	(0.02, 0.11)
0.10	0.10	0.01	0.06	(0.02, 0.11)
0.20	0.10	0.02	0.05	(0.01, 0.10)