

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Sociology

2021

THE DYNAMICS OF BEHAVIORAL ADOPTION: CULTURE, NETWORKS, INSTITUTIONS, AND THE SOCIAL GENOME

Sotoudeh, Ramina

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp015x21tj56n>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

**THE DYNAMICS OF BEHAVIORAL ADOPTION: CULTURE, NETWORKS,
INSTITUTIONS, AND THE SOCIAL GENOME**

Ramina Sotoudeh

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY
THE DEPARTMENT OF SOCIOLOGY

Advisor: Dalton Conley

September 2021

© Copyright by Ramina Sotoudeh, 2021. All rights reserved.

Abstract

Human behavior is complex, embedded in social and cultural structures and influenced by human biology. Across three different but interrelated studies, this dissertation integrates epistemological approaches to why and how people and groups adopt behaviors. The studies in this dissertation reveal how 1) cultural, 2) institutional 3) network-based and 4) biological mechanisms interact and jointly shape adolescent risky behavior. Study I (co-authored with Dalton Conley and Kathleen Mullan Harris) identifies metagenomic effects among adolescents, finding that having peers with higher genetic propensity to smoke leads adolescents to smoke more. In doing so, it reconceptualizes the social environment to include not only the behaviors of one's peers but their biologies as well. Study II assesses the relational consequences of institutional punishment and finds that smokers are more likely to cluster together in friendship networks when they attend schools that harshly punish smoking behavior. The extent to which clustering varies with punishment is moderated by one's genetic propensity to smoke. Those with lower genetic propensity to smoke are more likely to exhibit clustering based on smoking behavior in higher punishment environments. Higher smoking homophily and the harshness with which one's school punishes smoking are shown to be consequential for these adolescents' adult health outcomes. Finally, Study III argues for more holistic approaches to the study of human behavior. It reasons that any given behavior is best understood as part of a broader ecology of potential behaviors. This intuition is used to uncover multiple configurations of behaviors among American middle and high school students. It assesses the extent to which institutional setting, friendships and family environments explain these modalities of behavior and show that outcomes for drug users are conditioned by how drug use is embedded in the broader behavioral

ecology. Overall, the dissertation advances the development of an epistemologically integrated and multi-level approach to understanding behavioral complexity.

Acknowledgements

Many people helped make this dissertation a possibility. I am infinitely grateful to my dissertation chair, Dalton Conley, who taught me so much over the years. Thanks to Dalton, I was introduced to a whole new subfield which combined my passion for sociology and interest in biology. I am indebted to him for taking the time to teach me everything from the fundamental basics of biology, to complex methods, and research design, which all shaped the content of this dissertation. He influenced not only my knowledge, but my thinking; I am able to ask clearer questions, answer them more effectively, and create better research because of him.

I am also deeply indebted to Paul DiMaggio. Although we overlapped but a short semester at Princeton, Paul has been so influential in my transformation into a scholar. Working on our various projects over the years has been a constant source of joy and intellectual inspiration. I thank him for his generosity, for mentoring me from afar, for including me in exciting research projects, and for always making time to give insightful feedback.

I am also grateful to my other dissertation reader, Jennifer Jennings. I thank Jen for making the time to be on my committee, her valuable feedback, and her kindness and encouragement. I am thankful to Kathleen Mullan Harris, for being a part of my committee, for co-authoring the first paper in this dissertation, and for all the work she dedicated to the Add Health as the Principal Investigator over the many years; without a data set as rich as Add Health, this dissertation would not have been realized. I also owe a big thank you to Robert Wuthnow, who was a constant source of inspiration and whose door was always open to me over the years. I would also like to thank Cindy Gibson, Donna DeFrancisco and Amanda Rowe who were always so helpful and whose smiling faces I loved seeing in the pantry or in the halls. I

extend special thanks to Amanda, who I regularly troubled with all kinds of administrative issues while organizing my defense.

My journey to becoming a sociologist really started at NYU Abu Dhabi, where I had the pleasure of taking a research practicum course with Roger Friedland during my sophomore year. It was in that class that I discovered my love for sociology and the joy of doing research. Roger is one of the deepest thinkers and kindest people I know. He showed me how to be a good scholar, but more importantly to be beyond just that: to pursue what I truly enjoy and to treasure more than anything the beauties of friends, life, and family along the way. Thank you for your friendship and your continued support throughout my journey as a student of sociology.

At NYU Abu Dhabi, I was also fortunate to learn from Ivan Szelenyi, Hannah Brückner, Georgi Derluguian and Jui-Chung Allen Li. If I am lucky enough to teach sociology someday, I aspire to influence my students like they influenced me.

Writing this dissertation coincided with COVID-19 pandemic which made it a lonelier affair than it usually is. Nevertheless, I was lucky enough to have loving friends and family support me from near and far. I thank Arshia Sotoudeh, for being the most caring brother and for providing much-appreciated distractions from work when I needed them. I thank Mina Motamedi, Shirin Sadri, and Maral Kasiri for their love and support. I would not have been as efficient in writing, had I not had the weekends with them to look forward to. I thank Mastish Terefe for checking in on me from afar. Our conversations (be it about my dissertation or the ideal governance structure) are always a source of delight for me. I thank Cambria Naslund and Gelareh Farahi who were like my family in Princeton. I owe Gelareh a big thank you for getting me outside every week, whether to grab a drink, hike, or kayak, which helped wash away the stresses of being a graduate student. I am grateful to Parijat Chakrabarti, who has not only been a

great friend and officemate, but also my favorite person to exchange ideas with. The second paper in this dissertation especially benefited from conversations with him. I am grateful to Mark Hoffman, whose kindness and encouragement made writing this dissertation less difficult. I am thankful for our countless conversations and debates that continue to shape my thinking.

I would have accomplished nothing in life if not for my family. My parents are my role models in life, and I aspire to emulate their work ethic and adventurousness. I thank them for the many conversations we shared about the studies in this dissertation, and for their love and support every minute of every day, from near and far. Finally, I am thankful to my grandparents, Mamani and Babai, who have a special place in my heart, and at whose kitchen table in Tehran, I spent many hot summer afternoons writing. This dissertation is dedicated to them.

Table of Contents

Chapter 1: Introduction	1
Chapter 2: Effects of the Metagenomic Environment on Adolescent Smoking Behaviors	13
Chapter 3: Relational Consequences of Institutional Control	42
Chapter 4: The Embeddedness of Adolescent Risk Behaviors	88
Chapter 5: Conclusion	141
Supplementary Information for Chapter 2	144
Supplementary Information for Chapter 3	158
Supplementary Information for Chapter 4	169

Chapter 1

Introduction

Human behavior is complex. It is rooted in our multifaceted social life, biological systems, and their mutual interaction. Scholars from a variety of disciplinary backgrounds—from psychology to genetics to sociology—have explored how and why people come to develop and sustain behaviors. In varying ways, each approach and framework for understanding human behavior deals with adjudicating between two broad forces—selection and influence—which drive differences in behavioral adoption but are difficult to separate analytically. But the sources of selection and influence are many and their interactions innumerable. Because disciplines have tended to approach these questions on their own focusing, perhaps myopically, on candidate causes, the result is a disjointed understanding of how and why behavioral heterogeneity emerges.

Across three separate but interconnected papers, this dissertation integrates disparate epistemological approaches to why and how people and groups adopt behaviors. Each study in this dissertation is designed to understand how 1) cultural, 2) institutional 3) network-based and 4) biological mechanisms interact to jointly shape human behavior. Instead of tackling each of these etiologies separately, the research here concerns the co-constitutive relationship between human biology, culture, and institutions. I show that mechanisms within these various epistemological paradigms interact, amplifying or mitigating each other's effects, leading to unforeseen, emergent outcomes at the group-level. The implication is that only by studying these

mechanisms together can we come to understand why and how groups differ in the behaviors they adopt and sustain.

Where the papers differ in the mechanisms they examine, they share a substantive case: adolescent engagement in risky behaviors – specifically adolescent substance use. Substance use provides a strong case study because it is known to have both social and biological etiologies. Smoking addiction, for example, is known to be driven by clear biological pathways traceable to one's genes. And yet, variation in smoking behavior is widespread, even among those with the strongest genetic propensities to become addicted. As I will show, biological tendencies of the individual and group interact with cultural and institutional orientations towards smoking behavior as well as network-based mechanisms, which drive adoption and mediate identity formation. Behavioral adoption of smoking cannot be resolved to a single mechanism or epistemological etiology.

All three studies use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), which is a nationally representative data on adolescents' health and behaviors. There are many data sets collected on risky behaviors, but several factors make Add Health special. First, Add Health includes data from different domains, namely genetic and social data, which allows for the integration of these etiologies. But what makes Add Health different than other health surveys (some of which also include genetic data), is that social relations are embedded in institutional context and relational networks. The multi-level nature of the data makes it possible to view behavior as embedded in the context within which it happens, among other behaviors and in social relations. This enables me to take a more dynamic view of behaviors, to allow their meaning to vary and emerge relationally.

In what follows, I provide a substantive introduction to the case of adolescent substance use (including smoking and drug use) and make the case for an integrated approach to understanding them. I conclude by describing the three papers.

Background: Adolescent risky behavior and substance use

Adolescents live heavily institutionalized lives. Schools are not only the primary contexts in which adolescents pursue education, but also key sites for their personal development (Eccles & Roeser 2011), influencing their self-concept (Marsh et al 2008; Grieger et al 2014), social lives (Cronsoe et al 2003), experimentation with substances (Allison et al 1999), and experiences with institutional punishment and discipline (Skiba et al 2014). In schools, adolescents learn to be adults in a controlled environment; as such, schools are microcosms of the larger society into which adolescents will enter. Because schools are structured and relatively self-contained adolescent societies, they provide a good opportunity for researchers to study a multitude of dynamics at different levels of the social structure.

Habits developed in adolescence and in schools can last well into adulthood. Nine out of ten Americans who smoke cigarettes daily tried their first cigarette as a high school student (U.S. Department of Health and Human Services 2019). Experimentation with risky behaviors can have consequences: substance use in adolescence has been shown to be associated with a number of negative outcomes such as earlier school dropout (Gasper 2011; Chatterji & DeSimone 2005), and engagement with criminal justice system (Chassin et al 2016; Green et al. 2010). Higher adult morbidity rates have been reported for adolescents who smoke (Strand et al. 2004), as well as those who use illicit drugs (Johnstone et al. 2019). Adolescent smokers are more likely to also

engage in other drug use activities with additional comorbidities (Torabi et al. 1993), though the reasons behind this association remain contested (Bell & Keane 2014).

Understanding substance use and risky behaviors among adolescents has therefore long been a goal of social science and public health. Economists, sociologists, biologists, medical researchers, public health, and public policy researchers all study substance use and have independently forwarded theoretical paradigms for understanding their effects on health outcomes and why people engage in them. Some of these approaches have been integrated, while others are seen as competing and even contradictory. Perhaps the largest and longest standing epistemological divide in studying substance use (and human behavior more generally) is between social and biological perspectives. While biology is often seen as a characteristic of the individual, social explanations foreground the interaction between individuals.

Recently, these assumptions are being challenged. Rather than two distinct forces, scholars increasingly see nature and nurture as intertwined in complex and dynamic systems, feeding back into each other, permitting, enabling, and limiting agents' behaviors, mutually shaping outcomes, and producing heterogeneity across all different levels of human organization, from the individual to the collective.

Nature versus Nurture?

Since the sequencing of the human genome two decades ago, new advances in the field of genetics have made biological explanations more compatible with social ones. Geneticists are more and more acknowledging that genetic and social factors jointly shape outcomes, especially complex outcomes such as risky behaviors.

Unlike what many thought just years ago, there is no one gene that predicts a given behavior. Rather, most complex traits are “polygenic,” influenced by many genetic variants each with a small contribution, rather than monogenic (Boyle et al., 2017). This means that a single genetic variant has a very small impact on outcomes, and that the same gene might be involved in many different functions in the body. The growing evidence of polygenicity of many traits that social and behavioral scientists care about, in addition to the observation that the majority of variance in complex traits can be attributed to social components, is shifting our conception of genetic contributions to complex traits from one that is deterministic to one that is probabilistic. Which is to say, given the opportunity and exposure to a behavior, genes are now thought of as making it *more likely* for some individuals to adopt that behavior compared to others, but adoption is rarely guaranteed.

In recent years, large-scale gene discovery studies have found millions of genetic markers associated with complex traits, each of which has a miniscule effect on the trait. These small-but-many effects are summed into a score (a polygenic score) that can then be used in statistical models social scientists are familiar with. Gene discovery studies have found genetic loci related to many health and risky behaviors including smoking initiation and cigarettes per day (Lui et al 2019), number of lifetime sexual partners (Barban et al 2019), drinking behaviors (Walters et al. 2018), and a general risk tolerance score constructed using self-reported tendency for risk taking (Linner et al 2019). As a result, large social science data sets, including Add Health (Braudt and Harris, 2020), the Health and Retirement Study (Ware et al., 2018), the Wisconsin Longitudinal Study; the Fragile Families and Child Wellbeing Study and the Panel Study of Income Dynamics, have either already included or are in the process of including genetic data in addition to the standard social survey measures.

Beyond identifying straightforward genetic effects, however, the employment of polygenic scores in studies of social life have revealed important ways in which genes and the environment interact (Boardman et al. 2010; Schmitz and Conley 2016; Conley et al 2016; Barcellos et al. 2018, Johnson et al 2021). Increasingly, and for a range of outcomes, scholars in the field of sociogenomics are finding that the effects of genes on social outcomes depend crucially on social context. Work by Ryu and Lee (2019), for example, have found suggestive evidence that the different educational contexts in Korea and Europe may mean that different sets of genes predict educational outcomes. Similarly, in the Estonian context, genes played a different role in educational outcomes before versus after the fall of the Soviet Union (Rimfeld et al. 2018). Tropf et al. (2017), too, show that the role of genes in predicting social outcomes could differ as much as 75% across national contexts (this is compared to height as an outcome, a placebo case which we know is strongly determined by genes and whose heritability does not vary significantly across contexts). Relatedly, the effects of social contexts on outcomes are moderated by individuals' genes. For example, Barcellos, Carvalho, and Turley (2018) show that the effects of educational reforms in the UK in the 1970s on health outcomes were moderated by individuals' genetic propensity towards BMI. It is still early days for these gene-by-environment (GxE) studies but early evidence points to the idea that integrating perspectives from sociology, the sociology of culture, and sociogenomics may prove fruitful.

Integrating culture, networks and biology

How might integrating these perspectives provide new points of leverage for old problems? The three studies in the dissertation offer some different, but related, ideas. From the perspective of culture and networks, genes can improve how we attribute causation. Since cultural and

network/peer effects are largely endogenous forces, genes and institutions can serve as the exogenous leverage necessary for estimating causal effects. As I show in the first study of this dissertation (Chapter 2), because genes are determined at birth, when coupled with a design that eliminates the possibility of genetic selection (where individuals select into environments based on genetics), they can be used to overcome some of the most fundamental problems of estimating peer effects in social networks literature such as selection, contextual and reflection biases. In a related vein, we can leverage slight variations in their distributions across groups and populations, to model cascades of behavioral adoption, as long as the probability of adopting the behaviors in question has some genetic basis.

From the other perspective, by leveraging the complex endogenous processes of meaning making and network dynamics, we can come to understand the mechanisms through which exogenous forces like genes and institutions come to have their effects. For instance, prior research has shown that in an information environment where there is no common understanding of the link between tobacco use and disease, we observe low genotypic prediction of smoking behavior (Boardman et al. 2010). When the cognitive environment changed—after the Surgeon General’s 1964 report on the health risks of smoking (an institution-level change)—a strong genetic gradient to cigarette consumption emerges. The degree to which genes predict smoking behavior (or the heritability of smoking), therefore, can be thought of as mediated by the social and cultural context.

Integrating the study of culture and institutions with sociogenomics can help us better understand how and under what conditions genes have their effect; analogously, genes can help us better understand heterogeneous treatment effects, or why people appear to respond differently to the same policies or institutional environments. In Chapter 3, for example, I show that how students respond relationally to school-wide policies about punishing smoking behavior depends on their

genetic tendency to smoke. Policies written for one kind of student might have unintended consequences for others, which we will miss by focusing only on standard sets of individual-level variables.

Networks, culture, or even genes are not just properties of individuals, but they can be attributes of collectivities too. The first study (Chapter 2) shows how environments are not just sociocultural, but also metagenomic, or consisting of the genes of others, from classmates to friends. Networks and culture also offer tools for the study of collective behaviors. Since at least Breiger's seminal paper in 1974, "The Duality of Persons and Groups", sociologists have been applying network methods and a relational lens – that meaning inheres in the relationship between entities – to studying how individuals relate to culture, organizations, and behaviors. In doing so, they have identified *emergent collectivities*, groups of things (behaviors, organizations, objects) that come into association by virtue of their shared consumption or engagement. The third study in this dissertation (Chapter 4) uses relational methods to understand patterns in adolescent risky behaviors. Given that behaviors are both vertically as well as horizontally embedded, the third study uses Relational Class Analysis and Latent Class Analysis to discover distinct patterns of how behaviors covary in American middle and high schools. For some students, risky behaviors are isolated – their pattern of behaviors is characterized by an opposition between risky behaviors and social and school engagement. For others, risky behaviors happen alongside other engagements such as sports and home life. These configurations are nested, to varying degrees, in institutional (schools and grades), relational (friendship groups) and family environment; they are rooted in demographic characteristics, and can be explained, in no small part, with genetic factors. They are also associated with other important aspects of adolescents' lives: their happiness and likelihood to use drugs many years into the future.

Chapter Summaries

Overall, the papers in this dissertation collectively argue for a more integrated approach to studying human behavior. Chapter 2, co-authored with Kathleen Mullan Harris and Dalton Conley, asks whether the genes of those around us can affect our own outcomes. Such effects, what we term “metagenomic” effects, have been demonstrated for animals. However, in humans, where experimental manipulation is less feasible, such metagenomic effects of peers have largely been suggestive. We leverage as-if-random variation in same-grade peers to document metagenomic effects on adolescent smoking outcomes. The effect of grade-mates’ genotypes is larger in magnitude than many other predictors of smoking, including ego’s own genotype, being male, and family income. Further, a minority with high genetic propensities to smoke can affect the smoking behavior of an entire grade. This approach offers a way to integrate genetic and environmental influences on human phenotypes.

Chapter 3 seeks to understand how institutions impact social relations and the role that genes play in mediating the relationship between institutions and outcomes. Specifically, it examines whether institutional control, in the form of school punishment of smoking, impacts not only adolescents’ levels of smoking behavior, but the extent to which their social relations are homophilous with respect to smoking behavior. I find that schools with harsh punishment of smoking behavior exhibit increased smoking homophily. The tendency to cluster in high punishment environments is most common among the students with the lowest genetic propensity to smoke. Finally, I show that the homophily of one’s local network and the harshness with which one’s school punishes smoking together lead to more smoking behavior as an adult years after graduation.

Chapter 4 argues for an approach that considers adolescent risky behaviors as both vertically and horizontally embedded. In this study, I borrow methods commonly used in the sociology of culture (namely, Relational Class Analysis and Latent Class Analysis) to understand the patterns or modalities of behaviors that adolescents adopt across a range of behaviors from leisure activities to sports and club participation to risky behaviors like smoking, sex, and drug use. Three modalities of behaviors are revealed. In each risky behaviors relate to other behaviors in distinct ways: in one class risk-taking opposes all participatory activities in school including clubs, academics, and sports, and in two other classes risk-taking combines with sports and other forms of participation in school and home life. Students' membership in these behavioral classes is shaped by their institutional, relational, and familial commitments as well as their demographic backgrounds. Using twin models, I also estimate the heritability of class memberships, which proves to be roughly 30% across the classes. Finally, I assess the differences in drug users' happiness, and downstream drug use in the three relational classes. The results show that adolescent drug users who are also institutionally and socially embedded have better outcomes.

Reference

- Allison, K. W., Crawford, I., Leone, P. E., Trickett, E., Perez-Febles, A., Burton, L. M., & Le Blanc, R. (1999). Adolescent substance use: Preliminary examinations of school and neighborhood context. *American journal of community psychology*, 27(2), 111-141.
- Barban, N., Jansen, R., De Vlaming, R., Vaes, A., Mandemakers, J. J., Tropf, F. C., ... & Hyppönen, E. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature genetics*, 48(12), 1462-1472.
- Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences*, 115(42), E9765-E9772.
- Bell, K., & Keane, H. (2014). All gates lead to smoking: the 'gateway theory', e-cigarettes and the remaking of nicotine. *Social Science & Medicine*, 119, 45-52.
- Boardman, J. D., Blalock, C. L., & Pampel, F. C. (2010). Trends in the genetic influences on smoking. *Journal of health and social behavior*, 51(1), 108-123.
- Boardman, J. D., Blalock, C. L., & Pampel, F. C. (2010). Trends in the genetic influences on smoking. *Journal of health and social behavior*, 51(1), 108-123.
- Braudt, D., & Harris, K. M. (2020). Polygenic scores (pgss) in the national longitudinal study of adolescent to adult health (add health)—release 2.
- Breiger, R. L. (1974). The duality of persons and groups. *Social forces*, 53(2), 181-190.

- Chassin, L., Mansion, A. D., Nichter, B., & Pandika, D. (2016). Substance use and substance use disorders as risk factors for juvenile offending. In *APA handbook of psychology and juvenile justice*. (pp. 277-305). American Psychological Association.
- Chatterji, P., & DeSimone, J. S. (2005). Adolescent drinking and high school dropout.
- Conley, D., Laidley, T. M., Boardman, J. D., & Domingue, B. W. (2016). Changing polygenic penetrance on phenotypes in the 20 th century among adults in the us population. *Scientific reports*, 6(1), 1-5.
- Crosnoe, R., Cavanagh, S., & Elder, G. H., Jr. (2003). Adolescent friendships as academic resources. *Sociological Perspectives*, 46, 331-352.
- Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of research on adolescence*, 21(1), 225-241.
- Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of research on adolescence*, 21(1), 225-241.
- Gasper, J. (2011). Revisiting the relationship between adolescent drug use and high school dropout. *Journal of Drug Issues*, 41(4), 587-618.
- Green, K. M., Doherty, E. E., Stuart, E. A., & Ensminger, M. E. (2010). Does heavy adolescent marijuana use lead to criminal involvement in adulthood? Evidence from a multiwave longitudinal study of urban African Americans. *Drug and alcohol dependence*, 112(1-2), 117-125.
- Grieger, L., Kusunoki, Y., & Harding, D. J. (2014). The social contexts of adolescent romantic relationships. *Focus*, 31(1), 15-17.
- Johnston, L. D., Miech, R. A., O'Malley, P. M., Bachman, J. G., Schulenberg, J. E., & Patrick, M. E. (2019). Monitoring the Future National Survey Results on Drug Use, 1975-2018: Overview, Key Findings on Adolescent Drug Use. *Institute for Social Research*.
- Linnér, R. K., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., ... & Beauchamp, J. P. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature genetics*, 51(2), 245-257.
- Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., ... & Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature genetics*, 51(2), 237-244.
- Marsh, H. W., Trautwein, U., Lüdtke, O., & Brettschneider, W. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs Role of generalized and specific others. *Journal of Educational Psychology*, 100, 510-524.
- Rimfeld, K., Krapohl, E., Trzaskowski, M., Coleman, J. R., Selzam, S., Dale, P. S., ... & Plomin, R. (2018). Genetic influence on social outcomes during and after the Soviet era in Estonia. *Nature human behaviour*, 2(4), 269-275.
- Ryu, J., & Lee, C. (2019). Genome-Wide Association Study Reveals Two Nucleotide Variants Associated with Educational Attainment in Koreans. *Russian Journal of Genetics*, 55(9), 1130-1136.
- Schmitz, L., & Conley, D. (2016). The long-term consequences of Vietnam-era conscription and genotype on smoking behavior and health. *Behavior genetics*, 46(1), 43-58.
- Skiba, R. J., Arredondo, M. I., & Williams, N. T. (2014). More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity & Excellence in Education*, 47(4), 546-564.
- Strand, T. E., Malayeri, C., Eskonsipo, P. K., Grimsrud, T. K., Norstein, J., & Grotmol, T. (2004). Adolescent smoking and trends in lung cancer incidence among young adults in Norway 1954-1998. *Cancer Causes & Control*, 15(1), 27-33.
- Torabi, M. R., Bailey, W. J., & Majd-Jabbari, M. (1993). Cigarette smoking as a predictor of alcohol and other drug use by children and adolescents: evidence of the "gateway drug effect". *Journal of school health*, 63(7), 302-306.
- Tropf, F. C., Lee, S. H., Verweij, R. M., Stulp, G., Van Der Most, P. J., De Vlaming, R., ... & Mills, M. C. (2017). Hidden heritability due to heterogeneity across seven populations. *Nature human behaviour*, 1(10), 757-765.
- U.S. Department of Health and Human Services. Preventing Tobacco Use Among Youth and Young Adults: A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2012.
- Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., ... & Rice, J. P. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature neuroscience*, 21(12), 1656-1669.

Ware, E. B., Faul, J. D., Mitchell, C. M., & Bakulski, K. M. (2020). Considering the APOE locus in Alzheimer's disease polygenic scores in the Health and Retirement Study: a longitudinal panel study. *BMC Medical Genomics*, 13(1), 1-13.

Chapter 2

Effects of the Peer Metagenomic Environment on Smoking Behaviors

Co-authored with Kathleen Mullan Harris and Dalton Conley

Abstract

Can the genes of those around us affect our own outcomes? Research using family-based studies has shown that parents' uninherited genes also affect children's outcomes through social channels. We might expect these effects to also be operative outside of the family environment, since we know that peers heavily influence one another's behaviors. But in the absence of random assignment, the effect of other's genes on an individual's outcome is not easy to estimate. In this paper we leverage as-if-random variation in same-grade peers to document the effect of others' genes on adolescent's smoking outcomes – what we term “metagenomic effects.” The effect of grade-mates' genetic makeup is larger in magnitude than many other predictors of smoking, including an individual's own genetic tendency to smoke, being male and family income. Further, a minority with high genetic propensities to smoke can affect the smoking behavior of an entire grade. This approach offers a novel way to integrate genetic and environmental influences on human phenotypes. Overall, we argue that environments are not just sociocultural, but metagenomic, or consisting of the genes of others, from classmates and peers to parents and siblings.

Introduction

One of the most central aims of sociology as a discipline, is to understand how the social environment affects our beliefs, behaviors, and outcomes. Sociologists study a wide range of environments: from institutions, such as schools, to contexts such as neighborhoods, to the family environment and the friend group. In this study, we argue that we are all embedded in another, important environment: the environment made up of the genetic tendencies of those around us, or the metagenomic environment. Each of us live within social environments – groups of people who have distinct cultural frameworks, achievements, political beliefs, and, importantly, we argue, genetic tendencies. These genetic tendencies might become amplified in groups and impact outcomes of individuals around them regardless of an individual's own genetic tendency.

A growing body of research on animal genetics examines whether the genomes of other individuals can affect one's own phenotypes and has found that indeed the "metagenome" has a considerable impact on an individual's outcome. A recent experimental study on mice found that randomly assigned cage-mates' genotypes affect individual mice's outcomes, accounting for as much as 29% of the trait variance among the mice. In the absence of random assignment and environmental manipulation, these effects are hard to establish for humans. One recent study using Icelandic family-based genetic data, found that genes of parents, even those that are not inherited by the child, have a considerable impact on the child's outcomes through social pathways (Kong et al 2017). While Kong et al. show that parents genes impact individual outcomes above and beyond the direct transmission (vertical effects), how other, non-family-based (or horizontal) metagenomic environments impact us, is yet to be addressed. This is the task of the present paper.

To preview the findings, we show that the genetic propensity to smoke of one's peers affects one's own smoking behavior net of one's own genetic tendency and that this effect is likely

even larger than the individual's own genetic smoking risk. This is true not only when the peer group includes direct friends, and therefore where the individual plays an active role in shaping the metagenomic context, but also when the peer group includes all grade-mates (net of school level variation) and therefore when the individual does not select the metagenomic environment. Perhaps even more surprising is that a few people with high genetic risk to smoke can greatly affect the smoking behavior of an entire grade.

Adolescent Smoking

Tobacco use habits are largely established during adolescence; indeed, nine out of ten smokers first try cigarettes before age 18 (US Department of Health and Human Services 2014). Scholars have approached the question of which youth adopt smoking behavior from two perspectives. The first emphasizes individual's biological tendencies that reward nicotine consumption (Pidoplichko et al 1997) and in particular their genetic architecture that encodes that reward system (TAG Consortium 2010). Under this framework variation in tobacco usage and likelihood of developing a nicotine dependency result from individual genetic differences shaping the biological responses to nicotine consumption.

The second approach emphasizes instead the role the social environment plays in governing the availability of cigarettes (Woodruff et al 2003) and in acculturating individuals to smoke (Schofield et al 2003). For adolescents, the social environment may include many different contexts: the neighborhood (Mathur et al 2013), school (Perra et al 2012), their grade (Fletcher 2010), friends (Christakis and Fowler 2008), and family (Cambron et al 2018). Across these domains, same-age peers have been shown to be particularly influential: having coeval peers who smoke is associated with increased individual smoking (Nash et al 2005). This effect

holds for many different types of peers, with the literature showing a robust association between the smoking behavior of one's schoolmates, classmates, and close friends and one's own smoking behavior (Alexander et al 2001, Ennett and Bauman 1994). Limited experimental evidence suggests that these associations are not epiphenomenal (Gardner and Steinberg 2005).

Despite the respective power of genetics and peers in accounting for variation in which adolescents smoke, little prior research has sought to integrate the two—that is, to simultaneously identify genetic influence, peer effects and the interaction between the two as they influence adolescent tobacco use (Domingue et al 2015). One possible avenue for integration, and the one explored in this paper, is the analysis of metagenomic effects – also referred to as “social genetic effects” or “indirect genetic effects” in animal studies (Canario et al 2017; Bergsma et al 2008; Petfield et al 2005) —that is, the extent to which others' genes affect an individual's outcomes. We borrow the metaphor of metagenomics from research on the microbiome where all organisms in the microbiome are genotyped in order to capture the collective genomic context which is then used to understand individual outcomes. In the case here, the metagenome, or the collective genomic context, consists of the genomes of all of one's peers. Since smoking behavior has a genetic basis, and since others' behaviors may matter for individual behaviors and outcomes, we should not be surprised to observe metagenomic effects. Indeed, a recent experimental study on mice found that randomly assigned cage-mates' genotypes affect individual mice's outcomes, accounting for as much as 29% of the phenotypic variance among the mice (Baud et al. 2017).

However, for humans in particular, extant evidence for metagenomic effects is limited. Only a few studies to date have seriously considered the question and thus for most outcomes, including smoking, there is little credible evidence that such effects exist. Specifically, in the

absence of careful study design, genes and environments can act as mutual proxies, confounding the researcher's attempts to parse their effects (Domingue and Belsky 2017). Because of these difficulties, the few empirical studies that do consider the question of metagenomic in humans are, at best, suggestive (Raucher et al 2015; Cawley et al 2018). Perhaps the most convincing evidence for metagenomic effects in humans comes from research that has identified the important impact of parental and sibling genotypes on an individual's health and educational outcomes, in a dynamic called "genetic nurturance" (Kong et al. 2018; Conley et al. 2015). Outside the family environment, however, empirical evidence remains scarce; that said, research from behavioral genetics has consistently found that non-shared environment is more salient to most phenotypes than is the shared (read: family) environment of siblings (Polderman et al 2015; Polubriaginof et al 2018). This may suggest that metagenomics effects are not limited to genetic nurturance in the household context.

Analytic Approach

In this study, we explore the metagenomic effects of smoking. The data come from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a nationally representative cohort study of the health and behavior of U.S. students in grades seven through 12, which include genetic, social network and health data (Harris 2013). Add Health is a one-of-a-kind dataset, which ties adolescents and their genomes to their health and behaviors within the group setting most salient to their young lives: schools.

The basic analytic approach involves predicting an individual's smoking behavior as a function of the genetic risk for smoking in their metagenomic environment, operationalized as the mean polygenic score (PGS) for smoking of their peers, controlling for possible confounders

(such as population stratification) and the individual's own genetic risk (more detailed on this in the following section). In a second set of analyses, potential non-linear effects are explored by relaxing the mean polygenic score constraint.

By using unchanging peer genotypes as the independent variable, we are able to overcome two common problems in peer effects research: the reflection bias, where peer and individual behavior are observed at the same moment such that influence cannot be adjudicated, and the exclusion bias, where the fact that you cannot be friends with yourself induces a mechanical, negative correlation between peer and individual behavior (Manski 1993). That said, other potential biases thwart easy estimation. Peers and individuals share the same environment and therefore may adopt similar behaviors (contextual bias) (Cohen-Cole and Fletcher 2008; Christakis and Fowler 2013). For example, if people with particular genotypes attend similar kinds of schools, the estimation will be biased (Domingue et al 2018; Hamer and Sirota 2000). To ward against context effects, grade and school fixed effects are included, leveraging random, grade-cohort fluctuations in the genetic landscape to identify metagenomic effects.

Peers may also select into relations with one another, whether in being friends or classmates, on the basis of their genetic factors even within the context of a single grade. The risk that the results will be biased by this form of genetic selection will therefore depend on how peers are defined. Because adolescents are embedded in multiple social groups even within the context of a single grade, peers can be defined in a myriad of ways. Three such groups are explored here: the grade (all of the other students with whom one shares a grade, or grade-mates), the classroom (all of the other students with whom one shares classes, or classmates), and the friendship group (all of the friends one nominated or from whom one received a friendship nomination, or friends).

At the most macro-level, we explore the relationship between the average genetic predispositions towards smoking of an adolescent's grade-mates to the adolescent's own smoking behavior. We posit that conditional on school-level variation, the distribution of genetic propensity to smoke is as-if-random in a given grade, something we test using a simulation in the following section. Because genes are assigned at birth and being in a given grade within a school is roughly defined by the year and month of birth, it is reasonable to assume the absence of selection into the grade environment based on an individual's smoking genes. We show evidence for this assumption using a simulation design. Although smoking has been shown to be correlated with lower educational performance (Bryant et al 2000) and may therefore also be correlated in being held back a grade, there is no evidence of a nation-wide trend in individuals being held back in some years but not others. In a second set of analyses, we evaluate the extent to which the observed metagenomic effects are social in nature and operate through the actual smoking behavior of grade-mates using an instrumental variable analysis. While this analysis is likely underpowered, it suggests a novel way to link genetic and social theories, which could prove fruitful as more and larger datasets incorporate social, relational and genetic data.

We then explore the relationship between the average genetic predispositions of an adolescent's classmates and friends and the adolescent's smoking behavior. Unlike the grade-mate analyses, these models are descriptive in nature due to the presence of genetic homophily, but they are still analytically useful for contextualizing the grade-mate results presented earlier and are substantively interesting in themselves. The grade, while perhaps not considered the most salient peer group in the sociological literature, offers the chance to identify causal effects free from selection bias; effects of classroom peers and friends are subject to confounding due to selection; thus, the grade-mates effects can be considered the purest estimate of metagenomic

peer effects. The grade as an environment is also theoretically interesting because it is an ecology within which individuals in a birth cohort interact, form groups, and learn behaviors. Examining the entire grade relaxes the assumption that the diffusion of behaviors is restricted to dyadic interactions and that social contagion works like the contagion of viruses (Goldberg and Stein 2018; Crosnoe et al. 2004).

This framework also allows us to explore interesting and potentially superlinear effects of both peer behaviors and peer genes on individual behavior. For instance, we might expect individuals with the highest or lowest levels of genetic risk of smoking in a given grade to have a disproportionate impact on individual smoking compared to the impact of the average grade-level genetic risk. Research on the effects of grade and classroom composition on individual outcomes has pointed to the possibility that the presence of an exceptionally excellent student may lead to positive outcomes for everyone in the class (the “shining light” model) (Hoxby and Weingarth 2005). Alternatively, one or a few exceptionally disruptive students may evince negative outcomes for the entire grade/class (the “bad apple” model) (Lazear 2001; Sacerdote 2011). We test these contextual dynamics empirically by examining whether the proportion of individuals within a grade who fall in the top or bottom decile of the overall smoking polygenic score distribution, has a significantly different impact on individual smoking as compared to what the grade-level average PGS indicates.

Overall, our results show that adolescents’ smoking behavior is strongly predicted by the genetic propensity of their peers to smoke. This holds at multiple different levels, whether peers are conceptualized as grade-mates or friends. Further, adolescents appear to be particularly affected by peers who are in the top decile of smoking PGS, while peers in the bottom decile offer little by way of protection.

Data and Methods

The data for this study come from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a nationally representative cohort study on the health and behavior of adolescent school children first interviewed in 1994-95 across a sample of 132 middle and high schools across the United States (Harris et al 2010). An in-school survey was administered to every student present at each of the 132 schools asking them to self-report their social and health behaviors, friends, and family and school context. A follow-up in-home survey was administered to a random sample of the students in each school, which entailed a far more detailed questionnaire about their behaviors, attitudes, health, parents, siblings, and home life.

During the fourth wave of data collection in 2008-2009, 96% of the respondents who participated in the Wave IV in-home survey (N=15,159) agreed to provide biospecimen (in the form of saliva and capillary whole blood samples) to be immediately genotyped for specific SNPs and candidate genes; of those individuals, 80% (N=12,254) agreed to have their biospecimen archived for future use. With the more recent development of genotyping technology and the resulting decrease in cost, the archived samples were genotyped on about one million genetic markers, which provide genome-wide data for polygenic score construction (Harris et al 2013). Two Illumina platforms were used for genotyping Add Health genetic data. Illumina Human Omni1-Quad BeadChip was used for roughly 80% of the sample and includes over 1.1 million genetic markers. Illumina Human Omni-2.5 Quad BeadChip was used for the remainder of the sample and includes 2.5 million markers. A series of quality control procedures were performed on the SNP and individual level. Genetic markers with call rates $< 90\%$ and minor allele frequency $< 0.5\%$ were excluded. SNPs with Hardy-Weinberg Equilibrium P-value

$< 5 \times 10^{-5}$ were also excluded. After the quality control measures, genotype data were available for 9,975 individuals. For more details on the Add Health GWAS data quality see Highland et al. 2018 (Highland et al. 2018).

The measure of genetic propensity to smoke, common in the literature and one used in this paper, is a polygenic score. A polygenic score is a genome-wide score that summarizes the presence of presence and associated weights of risk alleles discovered by genome-wide association studies (GWASs). In essence, a polygenic score is a weighted average or composite score that considers information across an individual's entire genome to measure his/her genetic predisposition or risk for a particular outcome. A polygenic score for individual i is a weighted average across J SNPs of the number of reference alleles x (0,1 or 2) at loci j multiplied by the score for that SNP β :

$$(1) \quad PGS_i = \sum_{j=1}^J (\beta_j x_{ij})$$

Many complex traits have been shown to be highly polygenic, that is, affected by many different genetic variants each contributing a small amount to the overall outcome (Visscher et al 2008). This is true even for clinically dichotomous outcomes, which may reflect a shift along a phenotypic continuum known as decanalization (Gibson 2009). In analyzing such traits, individual genes will have low penetrance, making it difficult to distinguish between the effects of genes and the effects of the environment. Polygenic scores overcome many of these difficulties. Unlike candidate-gene approaches, polygenic scores do not focus on a few pre-specified genes of theoretical interest. Rather, they attempt to quantify an individual's genetic risk to exhibit a trait by aggregating the effects of all of genes on the phenotype of interest.

Polygenic scores are therefore “hypothesis-free”, in the sense the researcher does not have to understand all of the underlying biological processes giving rise to a phenotype to be able to study the effects of genes on that phenotype. As a result, they can be used to discover genetic contributions previously unknown to researchers, while correcting for false positive results that may result from multiple hypothesis testing.

Though polygenic scores are far more robust than candidate-genes, recent research has identified some of their shortcomings. Polygenic scores have been shown to pick up not only direct genetic effects, but also the effects of genetic nurturance, deep ancestry and assortative mating, all of which can bias results when the distribution of genes to environments is non-random. In order to assuage some of these concerns (specifically that ancestral genetic signals and cultural tendencies are correlated), researchers have used genetic principal components known to align with ancestry (Novembre et al. 2008). We also calculate the first four genetic principal components and residualize the polygenic scores on those components (essentially purging the polygenic score of the signal that is due to shared ancestry). That said, the individual-level biases of polygenic scores should not impact this paper’s results as long as the assumption holds that the distribution of grade-mate polygenic scores within schools is as-if random. Evidence in favor of this assumption is shown in the next section.

To identify bad apples and shining stars, respondents in the full sample were ordered according to their PGS. The top 10% of respondents in terms of PGS, those who were deemed as having the highest genetic risk of smoking, were classified as bad apples; while the bottom 10% were classified as shining stars. We performed this classificatory exercise separately for members of each race, in case racial differences in mean PGS were driving assignment to bad apple or shining star.

The polygenic score was constructed using summary statistics from the Tobacco and Genetics (TAG) Consortium's Genome-Wide Association Study (GWAS) on cigarettes per day (46). The genotyped variants in Add Health were then matched to the ones in the GWAS. A polygenic score was then constructed by aggregating the effects of the overlapping variants across the genome and weighting them by the strength of their association with the outcome or behavior. 529,390 single nucleotide polymorphisms (SNPs) were used to create the smoking scores.

Students who participated in the in-school administration of Add Health were asked to identify up to ten friends, five male and five female friends, using a roster of the other students in their school. The results from this exercise make up the friendship data used in this paper. As a consequence of the nomination design, every student is associated with as many as ten out-degree friends, and as many in-degree friends as there are other students in the school, though the large majority of students were nominated as a friend by others zero to four times. Friend-level variable estimates were obtained by averaging polygenic scores over: a) all of a student's friends; b) all of the people who nominated the student; or c) all of the people who the student nominated, which we refer to as friend-level, in-degree nominations, and out-degree nominations respectively. The focal individual is left out of the calculation of every peer-level calculation of the mean both for the independent variable (average smoking PGS levels) and for the peer-level control variables.

Identifying classmates required supplementary information, which came from the *The Adolescent Health and Academic Achievement (AHAA)* study. It provides the high school transcripts for Add Health Wave III sample members (original Wave I sample members who were re-interviewed at Wave III). After collecting the transcripts, the AHAA constructed

academic networks using overlap in course-taking and ran clustering analyses on those networks using a p^* -based algorithm fully detailed in Field et al. (2006). Individuals were assigned to clusters with other students with whom they took similar sets of courses, with the idea that the resulting clusters proxy mutual exposure during course taking. Students who were assigned to clusters with one or no other students were reassigned to the cluster to which they had the highest probability of belonging according to their coursework. These clusters were used to identify classmates, and again, for each student, averaged over the variables of their classmates to provide estimates of their peers' mean smoking behavior, mean smoking genetic risk, and demographic composition. We tested other specifications of classmates, including a simple binary indicator of having shared any course, all of which returned similar results as to those reported in the main text.

Individual smoking behavior measures how many cigarettes a day an individual smoked, on the days that they smoked. School and grade fixed effects are included to control for school and grade level. We further control for the sex, race, maternal education, familial smoking behavior, having an older siblings and family income of every individual in the sample as well as the average level of those variables at the grade level.

Further, following previous research on grade-level peers (Fletcher 2010), for all analyses, we limit the sample to students in schools with a 12th grade (which results in excluding middle schools, but retaining high schools with 7th and 8th grades) and who were assigned sample weights. Then, for all grade-level analyses, we further exclude students who attended a grade with fewer than 20 total students in our sample. In all, 74 schools and 148 grades are represented in the data. This leaves 3,895 respondents who have genotype data and are in grades with a

reasonable number of peers. For all friend analyses, only students who had at least one other friend in their grade were included, resulting in 3,708 respondents.

In addition to residualizing individual smoking PGS, a number of controls were included in the analyses in order to guard against various forms of population stratification (i.e. the correlation between shared genes due to population structure and shared environments). In terms of race, we include individuals of all races in the analyses, but control for the individual's own race and the racial composition of the grade, classroom, or friend group. Because it is possible that genes are a proxy for parent or sibling smoking behavior which might directly influence other individuals in a grade, we block this pathway by including an indicator for the presence of household smokers and older sibling both on the individual and on the grade level. In addition, we control for an individual's own smoking polygenic score in the models such that the results provide estimates of others' polygenic scores on individual smoking above and beyond an individual's own polygenic score.

Results

Simulation to evaluate as-if-random distribution of genes in grades

Before turning to the results, we evaluate the central assumption of this paper: namely that PGS is exogenous given grade and school fixed effects. This assumption amounts to the idea that, within a school, the assignment to grade is random with respect to genes and, specifically, smoking PGS. Another way to conceptualize this is that there is some gene pool for the specific school but selection into grade, in particular, is random, so that we can unbiasedly test between-grade differences.

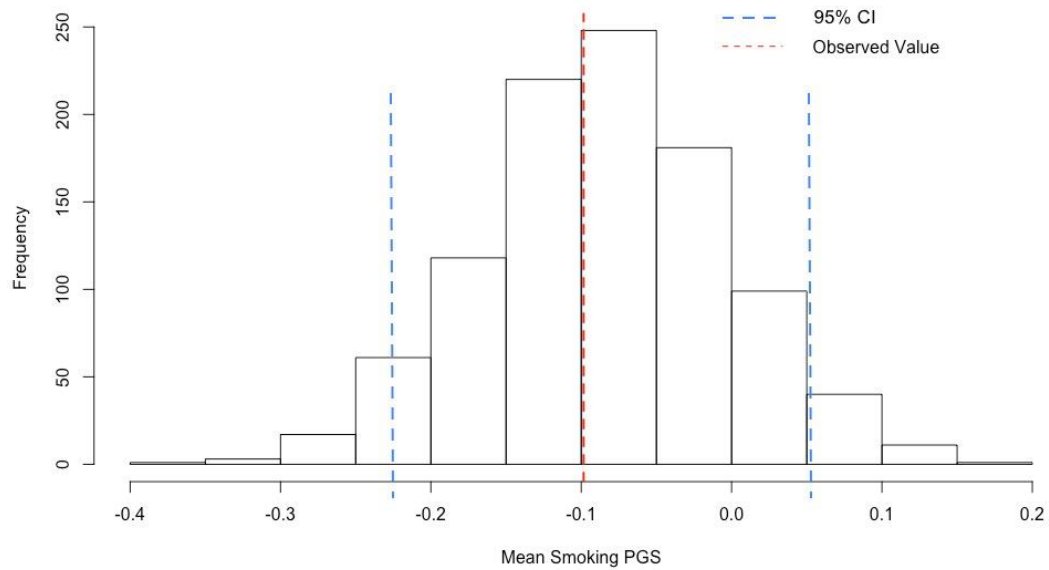
To test whether this is true, we ran a series of simulations where for each school, we randomly re-assign students' grades, holding the number of students in each grade constant, effectively jumbling the students' assignments to grade within each school, and as a result, the assignment of smoking PGS to grade. We do this 1000 times for each school, at each iteration calculating each grade's mean smoking PGS, giving each grade within a school 1000 bootstrapped mean smoking PGS scores for comparison.

We then test if the observed grade mean smoking PGS for each grade falls within the bounds of that grade's specific simulation results. To do this, we establish a 95% interval for each grade within each school by ordering the simulations for that grade according to their estimated mean smoking PGS value and set the bounds as the value at the upper and lower 2.5% of this ordered distribution. If, within a school, a grade's true smoking value falls within its simulated 95% interval, then we say that it is reasonable that it was drawn from the school superset at random.

Thus, for each school and grade combination, whether its observed mean smoking PGS value is reasonably random given the school PGS distribution, can be evaluated. We then calculate what proportion of school-grade combinations appear randomly drawn. If the assumption that grade PGS are drawn at random from school PGS is true, this proportion should be close to 95% (i.e. the size of the interval).

The results (visualized in Figure 1) show that 96.2% of grades fall within 95% bounds of their simulation, giving us confidence that the central assumption holds.

Figure 1: Visual inspection of Simulation Results



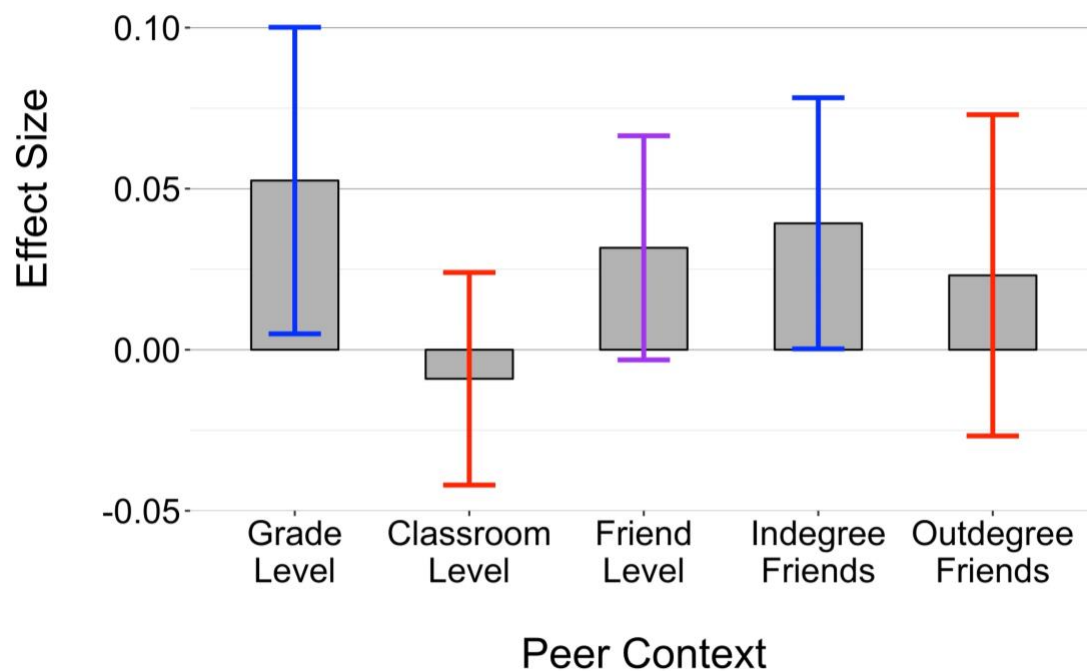
Metagenomic effects

In the results presented in Figure 2, the association between the metagenomic environment, defined as grade-mates', classmates', and friends' smoking polygenic scores, and individual smoking behavior is assessed. Bars signify the magnitude of the coefficient for each variable. CIs at the 95% are colored according to significance: blue if significant at a 0.05 level, purple if significant at a 0.10 level, and red otherwise.

To re-iterate, grade-mates include all others in an adolescent's grade, classmates include people with whom the adolescent commonly shared classes, and friends include any student who the adolescent nominated as a friend and anyone who nominated the adolescent as a friend. For each model, controls for race and the smoking behavior of family members, in order to ward against potential confounds related to population stratification, and a number of contextual variables shown to be associated with smoking behavior, both on the individual and respective

peer level, were included. That is to say, models at the grade-level control for the attributes and behaviors of the grade-mate peers, while models at the friend-level control for those very same attributes but averaged across a person's friend group. When considering peers at a more macro-level such as the grade, friend or classroom clusters were not excluded because we are interested in the overall effect of the metagenomic environment and not the effect of the grade net of the friends and shared classroom contexts. Recall that the focal individual is left out of the calculation of peer-level measures and, thus, unlike a classical hierarchical model, the number of observations in each regression reflects the total number of students in Add Health who have membership in a grade, classroom or friendship group, rather than the number of grades, classrooms or friendship groups in the data. All models correct standard errors for school-grade clusters.

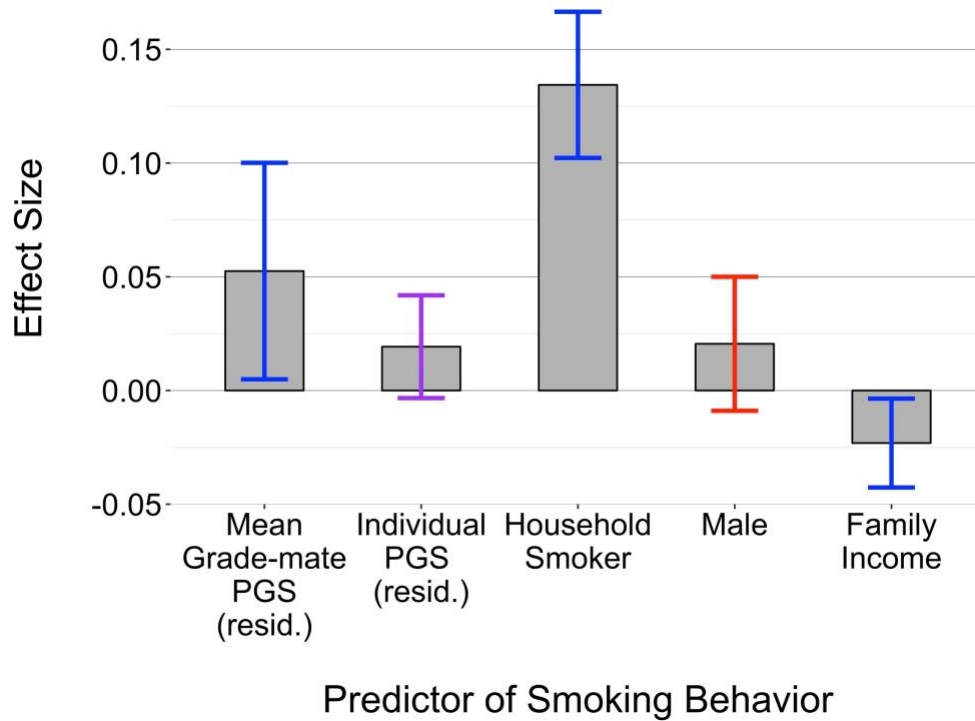
Figure 2: Metagenomic Effects in Different Peer Contexts



For three of the five models, we find that the average levels of smoking polygenic scores among peers is positively and, depending on the model, at least marginally, significantly associated with individual smoking behavior. This association is positive and significant at the grade-level, which is the only causally identified model. It is marginally significant for friends. That the confidence intervals of the grade and friend-level models are overlapping suggests they do not significantly differ from one another. That said, we would expect the friend estimates to be higher than grade since individuals select friends who are alike them, and are influenced by their friends. In the Appendix, we show that this is likely due to the small number of friends sampled in Add Health.

Breaking down friendships by indegree friends (those who nominated an individual) and outdegree friends (those whom an individual nominated) shows that indegree friends' average smoking PGS is positively and significantly associated with smoking behavior while outdegree smoking PGS is not. Classmates' smoking PGS is not significantly associated with individual smoking outcomes, likely indicating that classmates are not a highly relevant social context and that classmates are inconsequential for one's smoking behavior.

Figure 3: Comparing the effect grade-mate smoking genetic risk to known individual-level predictors on smoking.



In Figure 3 the grade-level metagenomic effect alongside other coefficients included in the grade-level model, are plotted. Since all the coefficients come from the same model, the effects are conditional on each other. These factors have been identified as important determinates of adolescent smoking, including: individual smoking polygenic score (Domingue et al 2015), being male (Grunberg et al 1991), having a household member that smokes (Fletcher 2010), and family income (Winkleby et al 1992). The sizable metagenomic effect represents a form of genetic nurturance that is “horizontal” in nature, in contrast to the vertical nurturance parental genetic nurturance documented by Kong et al. (2018) and which may be driving the strong effect that having a household member that smokes has on one’s smoking behavior.

Instrumental Variable Analysis

The grade-mates' genetic tendency in this research design allows us to answer age-old questions about peer effects. The study of peer effects, that is, the causal effect of peers on individuals' behaviors, beliefs or outcomes, suffers from a set of difficulties that thwart its estimation: 1) the reflection problem, where it is unclear whether peers are influencing the individual or vice versa; 2) contextual bias, where a common factor may be leading to the observed similarity between two peers, 3) the selection problem (or homophily), where individuals with similar attributes or genotypes tend to become friends, and 4) exclusion bias, or the fact that one cannot be friends with oneself, which constrains peer mixing patterns and induces mechanical negative correlation between individual and peer behavior. Because genes are assigned at birth and being in a given grade is roughly defined by one's age, we have argued, and provided some evidence using a simulation, that the distribution of genetic propensity to smoke in a given grade is as-if random conditional on the school attended. This provides a plausible identification strategy for the estimation of the effect of peer smoking on individual smoking.

To estimate the peer effects of smoking behavior, we use the average level of genetic predisposition toward smoking within a school and grade as an instrument for the levels of smoking behavior in that grade. Using two-stage least squares model we are able to quantify the effect of peer polygenic score on individual smoking behavior as it operates through peer behavior.

The first stage of the two stage least square is the following:

$$(1) \quad \bar{Y}_{-igs} = \alpha + \gamma_1 \bar{Z}_{-igs} + \gamma_2 G_{igs} + \gamma_3 W + \xi_{igs}$$

where igs indexes an individual i in grade g in school s and $-igs$ signifies that the estimate excludes the individual. Y is smoking behavior of the individual, $\bar{Y}_{-igs}$ is the average smoking level of peers excluding the individual, G is the smoking polygenic score for a given individual, $\bar{G}_{-igs}$ is the average level of the smoking polygenic score within a school and grade excluding the individual and W is the vector of controls (including individual level and grade level variables). The additive error term is ξ .

Equation 2 portrays the second stage regression of our model:

$$(2) \quad Y_{igs} = \alpha + \rho_1 \hat{\bar{Y}}_{-igs} + \rho_2 G_{igs} + \rho_3 W + v_{igs}$$

where $\hat{\bar{Y}}_{-igs}$ indexes the fitted values from the first stage regression, and v is the error term.

Using this model, the peer effect of smoking as instrumented by the genes of peers can be identified. Beyond helping with causal identification, this model also reveals a substantively interesting quantity: a point estimate of the impact of one's peers genetic risk of smoking on an individual's smoking as it operates through peer behavior. The results can be found in Figure 4 (corresponding tables can be found in Appendix A). and they show a positive and significant relationship between peer and individual smoking, with a point estimate similar to that of the metagenomic effects models (presented in Figure 2), signaling that the effect of peer genes operates through behavior.

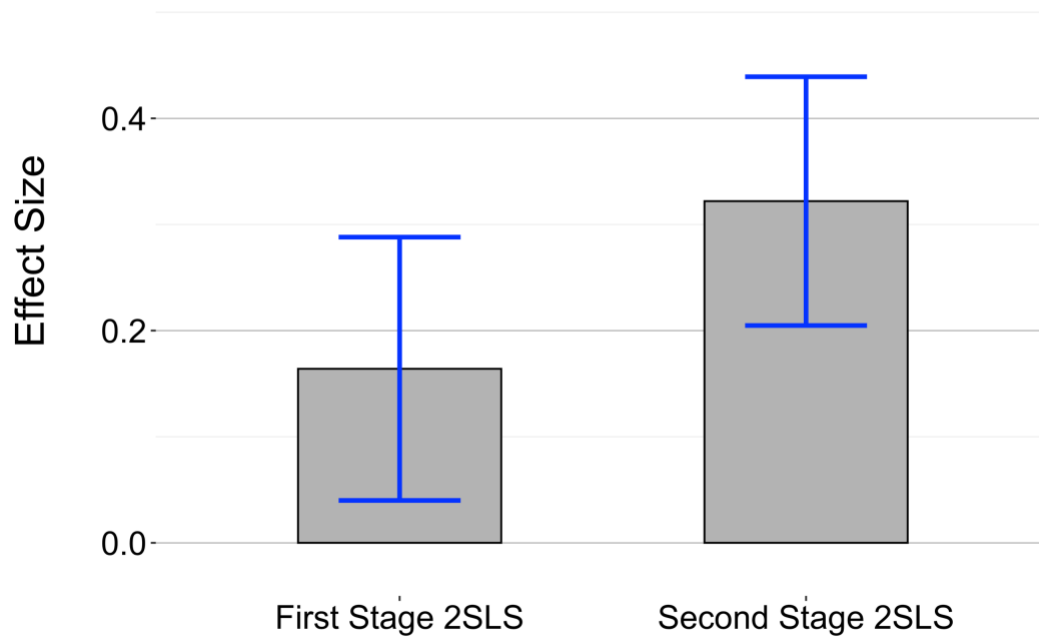
Like any other instrumental variable analysis, we make a series of assumptions, including that the instrument is sufficiently strong in the first state, monotonicity (that there are not individuals for whom having peers that smoke more make them smoke less), and that the exclusion restriction assumption is not violated (peer PGS does not affect ego smoking other than through the smoking behavior of peers). A series of analyses were undertaken to test

whether these assumptions are reasonably true. One main concern is pleiotropy: where the same genes contributing to smoking polygenic score also have an impact on say, a person's friendliness or agreeableness and therefore their influence on more people's behaviors.

To evaluate whether this was the case, we performed simple correlation analyses, where we correlated grade-level network density with the grade-level mean polygenic score for smoking, produced by residualizing each individual's polygenic score on their first four PCs and then averaging over each grade. The resulting grade-level polygenic scores were further residualized by school fixed effects before estimating the correlation. This returned a very low correlation (0.03), which was insignificant at 0.05 level. we performed similar analyses at the individual (ego-centric) level and found similarly trivial and insignificant correlations. Though the possibility of unaccounted pleiotropy (i.e. an exclusion restriction violation) cannot be completely dismissed, the correlational exercise rules out obvious hypotheses regarding some potential pleiotropic effects of smoking risk and network measures.

A series of tests were also undertaken to ensure the strength and robustness of the instrument. A Lagrange-Multiplier (LM) test for under-identification using the Kleibergen and Paap (2006) rk statistic was used to test that the instrument was not under-identified, while an F statistic was used to test that the instrument was not weak. Because we are using clustered standard errors, we cannot use the standard F-statistic of the first stage as a test of weak or strong instrument. The null was rejected for the under-identification test. A traditional Cragg-Donald Wald test returns an F-statistic of 217.5, while a Kleibergen-Paap rk Wald F statistic, which is more robust when i.i.d is not met, returns an F-statistic of 6.7. This discrepancy signals that the instrument may be weak, so the results should be interpreted cautiously.

Figure 4: Metagenomic effects as instrumental variable



Superlinear metagenomic effects

The models above have shown an association of higher peer smoking genetic risk and individual smoking outcomes. This effect, however, might not be strictly linear. Individuals who have the highest genetic predisposition to smoking could have a disproportionate impact on an individual's smoking. We examine this possibility by replacing the average level of genetic risk of smoking by the proportion of the grade that has a top or bottom decile smoking PGS (see Materials and Methods for how the deciles were calculated), which we refer to as “bad apples” and “shining stars” models, respectively, following previous literature (Hoxby and Weingarth 2005).

The results can be found in Figure 3, where Panel A presents a coefficient plot for the bad apple regressions and Panel B presents a coefficient plot for the shining star regressions. Proportion bad apples is positively associated with one's own smoking across all of the different operationalizations of the peer group except for the classroom, but only significantly so at the grade-level. Proportion shining stars is negatively associated with one's own smoking across all of the different operationalizations of the peer group except for the classroom, but never significantly so. Therefore, ignoring significance levels the direction of the coefficients, the results generally paint a similar picture across all levels except for the classroom: higher proportions of bad apples are positively associated with smoking behavior, while higher proportions of shining stars do not appear to have a suppressive effect.

Figure 3: Metagenomic effects of top decile (bad apples) and bottom decile (shining stars) smoking PGS on adolescent smoking outcomes

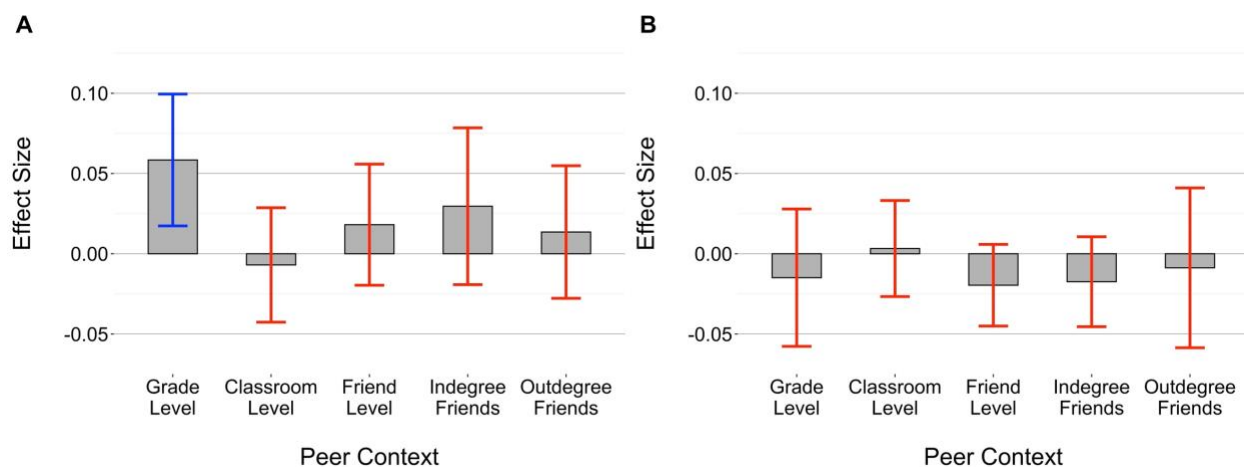


Figure 3. This figure shows the effect of bad apples and shining stars on smoking behavior across different peer contexts. A compares the effect of bad apples across different peer contexts, while B does the same for shining stars. Bars signify the magnitude of the coefficient for each context. CIs at the 95% are colored according to significance: blue if significant at a 0.05 level, purple if significant at a 0.10 level, and red otherwise.

Conclusion

Smoking behavior has long been thought to have both social and genetic etiologies. This paper suggest that genes should be treated as an important part of the social environment and the social environment must be considered when thinking about genetic influence. In this paper, the metagenomic environment is shown to be an important factor for individual outcomes: metagenomic environments characterized by higher smoking tendencies, lead individuals to smoke more. These associations are observed across multiple social contexts: the grade, the friendship group and in-degree friends (those who nominated the individual as a friend). Grade-mate smoking behavior is a stronger predictor of individual smoking than even the individual's own genetic contributions (as the magnitudes of the variables show).

We also explored potential superlinear metagenomic effects and find that the presence of “bad apples” – those with exceptionally high genetic risk of smoking – in the peer context, increases adolescents' smoking likelihoods. More work should be done to understand under what conditions the effect of bad apples can be mitigated (or, for that matter, the effect of shining stars can be amplified), including how forces such as the grade friendship network structure, school's administrative activities, or state's cigarette taxation rates might moderate their effect.

The theoretical contributions of the paper are threefold: first, it contributes to sociogenomics literature by showing that genes, whose only task is to encode for proteins in one individual's body, impact others around us through social processes. Similar findings have been established in animal studies where individuals are shown to be impacted by other individuals' genetic tendencies, but in the absence of random assignment to different genetic environments, this had not previously been established in humans, and this is the first such study to do so.

From a sociological point of view, thinking about the social aspects of genetics leads us to reconceptualize the social environment. Our social environment is not just made up of the observable characteristics of our peers (like their race, class, gender, educational achievements), but also their predispositions and tendencies, some of which are genetically endowed. Taking account these latent collective predispositions will help us understand why behaviors are taken up by some groups and not others, and why they persist.

In a similar way, it gives the concept of ‘preferences’ a locus of measurement – much social science research is oriented towards deciphering where preferences come from, and how to measure them. Economists often argue that preferences are revealed through behavior, whereas sociologists use declared preferences or attitudes measured by surveys. Genes give us another avenue through which to gauge potential preferences.

And finally, genes are inherited at birth, and can help us solve the chicken and egg problem with peer effects. We can use the genes of other people as something that makes them more likely (even just slightly) to engage in a behavior – a sort of intent-to-treat – and we can then watch as that behavior spreads through a network. Genes, coupled with the analytical design of this paper which only relies on random variation in the local metagenomic environment, future research can estimate both metagenomic effects and peer effects for any behavior that has a genetic basis. This is especially valuable given the difficulty of estimating causal effects of peers, whether they result from their behaviors or genes.

References

- Alexander, C., Piazza, M., Mekos, D., & Valente, T. (2001) Peers, schools, and adolescent cigarette smoking. *J Adolesc Health* 29: 22-30.
- Baud A., et al. (2017). Genetic Variation in the Social Environment Contributes to Health and Disease. *PLoS Genet* 13: e1006498.
- Bergsma, R., Kanis, E., Knol, E. F., & Bijma, P. (2008). The contribution of social effects to heritable variation in finishing traits of domestic pigs (*Sus scrofa*). *Genetics* 178:1559-1570.
- Bryant, A.L., Schulenberg, J., Bachman, J.G., O'malley, P.M., & Johnston, L.D. (2000). Understanding the links among school misbehavior, academic achievement, and cigarette use: A national panel study of adolescents. *Prev Sci* 1: 71-87.
- Cambron, C., Kosterman, R., Catalano, R.F., Guttmanova, K., & Hawkins, J.D. (2018). Neighborhood, Family, and Peer Factors Associated with Early Adolescent Smoking and Alcohol Use. *J Youth Adolesc* 47: 369-382.
- Canario, L., Lundeheim, N., & Bijma, P. (2017). The early-life environment of a pig shapes the phenotypes of its social partners in adulthood. *Heredity* 118: 534.
- Cawley, J., Han, E., Kim, J.J., & Norton, E.C. (2018). Testing for Peer Effects Using Genetic Data (No. w23719). *Natl Bur Econ Res*.
- Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: examining dynamic social networks and human behavior. *Stat Med* 32: 556-577.
- Christakis, N.A. & Fowler, J.H. (2008). The collective dynamics of smoking in a large social network. *New Engl J Med* 358: 2249-58.
- Cohen-Cole, E., and Fletcher, J. M. (2008). Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *J Health Econ* 27: 1382-1387.
- Conley, D., Domingue, B. W., Cesarini, D., Dawes, C., Rietveld, C. A., & Boardman, J. D. (2015). Is the effect of parental education on offspring biased or moderated by genotype?. *Sociological science*, 2, 82.
- Crosnoe, R., Muller, C., & Frank, K. (2004). Peer context and the consequences of adolescent drinking. *Soc Probl* 51: 288-304.
- Domingue, B. W., & Belsky, D. W. (2017). The social genome: Current findings and implications for the study of human genetics. *PLoS Genet* 13: e1006615.
- Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M., & Boardman, J. D. (2015). Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health. *AERA open* 1: 2332858415599972.
- Domingue, B. W., Belsky, D. W., Fletcher, J. M., Conley, D., Boardman, J. D., & Harris, K. M. (2018). The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health. *Proceedings of the National Academy of Sciences*, 115(4), 702-707.
- Ennett, S.T. & Bauman, K.E. (1994). The contribution of influence and selection to adolescent peer group homogeneity: the case of adolescent cigarette smoking. *J Pers Soc Psychol* 67: 653.
- Field, S., Frank, K.A., Schiller, K., Riegle-Crumb, C., & Muller, C. (2006). Identifying Social Contexts in Affiliation Networks: Preserving the Duality of People and Events. *Social Networks* 28: 97-123.
- Fletcher, J.M. (2010). Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics* 19: 466-484.
- Fletcher, J.M. (2010). Social interactions and smoking: evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Econ* 19: 466-84.
- Furberg, H., et al.; Tobacco and Genetics Consortium (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 42: 441-447.
- Gardner, M. & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Dev Psychol* 41: 625-35.
- Gibson, G. (2009). Decanalization and the origin of complex disease. *Nature Reviews Genetics*, 10(2), 134.
- Goldberg, A., & Stein, S. K. (2018). Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation. *Am Sociol Rev* 83: 897-932.
- Grunberg, N. E., Winders, S. E., & Wewers, M. E. (1991). Gender differences in tobacco use. *Health Psychol* 10: 143.
- Hamer, D., & Sirota, L. (2000). Beware the chopsticks gene. *Mol Psychiatry* 5: 11-13.

- Harris, K. M., Halpern, C. T., Hussey, J., Whitsel, E. A., Killea-Jones, L., Tabor, J., ... & Smolen, A. (2013). Social, behavioral, and genetic linkages from adolescence into adulthood. *American Journal of Public Health*, 103(S1), S25-S32.
- Harris, K.M. (2010). An integrative approach to health. *Demography* 47: 1-22.
- Harris, K.M. (2013). The add health study: Design and accomplishments. Chapel Hill: Carolina Population Center, UNC at Chapel Hill.
- Highland, H.M., Avery, C.L., Duan, Q., Li, Y., & Harris, K.M. (2018). Quality control analysis of Add Health GWAS data. URL: http://www.cpc.unc.edu/projects/addhealth/documentation/guides/copy_of_AH_GWAS_QC.pdf.
- Hoxby, C.M. & Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Working Paper No. 7867.
- Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsen, B. J., Young, A. I., Thorgeirsson, T. E., ... & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359(6374), 424-428.
- Lazear, E.P. (2001). Educational production. *Q J Econ* 116: 777-803.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Rev Econ Stud* 60: 531-542.
- Mathur, C., Erickson, D.J., Stigler, M.H., Forster, J.L., & Finnegan, J.R. (2013). Individual and neighborhood socioeconomic status effects on adolescent smoking: a multilevel cohort-sequential latent growth analysis. *Am J Public Health*. 103: 543-8.
- Nash, S.G., McQueen, A., & Bray, J.H. (2005). Pathways to adolescent alcohol use: family environment, peer influence, and parental expectations. *J Adolesc Health* 37: 19-28.
- Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., ... & Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature*, 456(7218), 98-101.
- Perra, O., Fletcher, A., Bonell, C., Higgins, K., & Mccrystal, P. (2012). School-related predictors of smoking, drinking and drug use: evidence from the Belfast Youth Development Study. *J Adolesc* 35: 315-24.
- Petfield, D., Chenoweth, S. F., Rundle, H. D., & Blows, M. W. (2005). Genetic variance in female condition predicts indirect genetic variance in male sexual display traits. *Proc Natl Acad Sci USA* 102: 6045-6050.
- Pidoplichko, V.I., Debiase, M., Williams, J.T., & Dani, J.A. (1997). Nicotine activates and desensitizes midbrain dopamine neurons. *Nature* 390: 401-4.
- Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics*, 47(7), 702-709.
- Polubriaginof, F. C., Vanguri, R., Quinnes, K., Belbin, G. M., Yahi, A., Salmasian, H., ... & Tatonetti, N. P. (2018). Disease heritability inferred from familial relationships reported in medical records. *Cell*, 173(7), 1692-1704.
- Rauscher E., Conley, D., & Siegal, M.L. (2015). Sibling genes as environment: Sibling dopamine genotypes and adolescent health support frequency dependent selection. *Soc Sci Res* 54: 209-20.
- Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*.
- Schofield, P.E., Pattison, P.E., Hill, D.J., & Borland, R. (2003). Youth culture and smoking: Integrating social group processes and individual cognitive processes in a model of health-related behaviours. *J Health Psychol* 8: 291-306.
- Tobacco and Genetics Consortium (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 42: 441-7.
- U.S. Department of Health and Human Services. (2014). The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
- Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nature reviews genetics*, 9(4), 255.
- Winkleby, M. A., Jatulis, D. E., Frank, E., & Fortmann, S. P. (1992). Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health* 82: 816-820.
- Woodruff, S.I., Candelaria, J.I., Laniado-Laborin, R., Sallis, J.F., & Villaseñor, A. (2003). Availability of cigarettes as a risk factor for trial smoking in adolescents. *Am J Health Behav* 27: 84-8.

Chapter 3

Relational Consequences of Institutional Control

Abstract

Sociologists increasingly find that institutions shape social networks through a range of mechanisms. Though institutions vary in their purposes and structures, many share a tendency to use punishment as means of social control. However, few studies consider how punishment relates to network formation. Using the National Longitudinal Survey of Adolescent to Adult Health, I examine the effects of institutional control, in the form of the punishment of smoking behavior, on adolescent friendship networks and the extent to which smoking is a basis of friendship formation. I show that a) schools with harsh punishment of smoking behavior exhibit increased smoking homophily, b) those with lower genetic propensity to smoke are more likely to exhibit clustering based on smoking behavior in higher punishment environments, and c) that the homophily of one's local network and the harshness with which one's school punishes smoking, together lead to more smoking behavior as an adult, many years into the future.

Introduction

This study considers the relational consequences of institutional control, in the form of the punishment of smoking behavior in American high schools. A rich literature documents the consequences of punishment on punished individuals, whether with respect to child rearing (Haapasalo & Pokela 1999; Gershoff 2010), in schools (Perry & Morris 2014; Kupchik 2016), or criminal justice outcomes (Bernburg & Krohn 2003; Lopes et al. 2003; Chiricos et al 2007).

Drawing on labeling (Lemert 1967) and stigma theories (Goffman 1963), studies have found that punishment in schools affects adolescents through a number of distinct mechanisms including shaping their self-concepts as law breaking individuals (Matsuda 1992; Schwartz & Skolnick 1962), reducing their access to opportunities (Bernburg and Krohn 2003; Bushway 1998; Lanctôt et al. 2007), and increasing their identification with other deviant individuals (Bernburg et al. 2006; Ascani 2012) with near-term and downstream consequences (Mallett 2016).

Externalities of institutional control, however, extend beyond the individual. Institutions shape the relational contexts in which individuals are embedded. The literature on adolescent networks has shown, for example, that friendship formation in schools is influenced by a number of institutional factors, such as school size, how they track their students, whether their policies are tolerant, and how they manage their classrooms (McFarland et al. 2014; McNeely et al. 2002).

While we know generally that institutions, and schools in particular, shape social relations, we know less about the specifics. Schools have a number of different tools at their disposal for shaping the social relations of their students from incentive structures that reward good behavior to rules about who and where students can associate with one another, to the harsh punishment of inappropriate or unwanted behaviors (Hoffman et al. 2009). Each of these tools shapes social relations in unique ways and through distinct mechanisms. As such, it is important to consider forms of control carefully to understand their distinct relational effects.

In addition, a burgeoning literature at the intersection of sociology and genetics finds evidence that genes influence how individuals respond to institutional contexts. For example, genetic factors have been shown to play an important role in moderating individual responses to institutional changes in policy, both in education (Barcellos et al. 2018; Johnson et al. 2020) and cigarette consumption (Fletcher 2012). Given that genes also appear to play a role in friendship

formation (Fowler & Christakis 2011; Domingue et al. 2018) and spousal selection (Domingue et al. 2014), we might expect important interactions between genes and institutions in the formation of relationships.

With these concerns in mind, this paper considers how punishment of smoking in schools shapes social relations with respect to cigarette consumption, and the extent to which the relational effects of school punishment are moderated by individuals' genes. Where most previous literature on the institutional forces that shape student networks, has focused on either incentives or forms of tracking in schools, I focus on punishment, which is important both for its pervasiveness (nearly every school in the United States uses punitive measures to affect student behavior) and for its broader entanglements with criminal justice and societal control (Cuellar & Markowitz 2015). Punishment marks students as deviant, reinforces existing inequalities, and influences negative educational outcomes (Lemert 1967; Kupchick 2016). Its relational consequences may be similarly significant, and may in fact be mechanisms, contributing to inequalities in outcomes (DiMaggio and Garip 2014; Centola 2011).

Using data from National Longitudinal Study of Adolescent to Adult Health (Add Health), a nationally representative study of students' health behaviors, friendship patterns, and school environments, I leverage school-level, institutional variation in the harshness with which schools punish smoking to understand how punishment influences the patternings of social relations. Since at least 1988, nearly every American high school has had an explicit, written policy punishing smoking on school grounds (CDC 1989). However, there is a great deal of variance—especially compared to more serious offenses such as being under the influence of alcohol or illicit drugs at school—in the harshness with which schools punish smoking: some levy verbal warnings for the first offense, while others suspend students caught smoking on school grounds.

I focus on the extent to which harshness in smoking punishment shapes the patterning of friendship patterns: does punishment influence who students befriend with respect to smoking status? There are good reasons to expect so, though the expected direction of the effect is unclear. In highly punitive environments, for example, students who smoke may increasingly rely on each other to support their habit and avoid detection. They may also increasingly identify with one another in contexts where punishment renders smoking more socially salient. On the other hand, smokers may be more likely to hide their smoking punishment when it is punished more harshly, leading smoking to become a more private and individualized behavior as opposed to a social marker. These different mechanisms may in turn be moderated by individual attributes including genetic tendency to smoke. In recent waves, Add Health has added information about the genes of respondents, allowing me to examine whether and how punishment affects the friendship formation of those who are more at risk of smoking addiction. Lastly, I ask whether these institutional contexts and relational patterns matter for later life outcomes.

The results indicate that smoking punishment has a strong influence on the patterning of social relations: individuals who smoke are more likely to have other smokers as friends when smoking punishment is harsh. These effects are strongest when students are in ninth, tenth and eleventh grades, but disappear when they enter twelfth grade, when smoking becomes, for most students, legal if done off school grounds and schools become much less likely to punish students in general. These effects are moderated by the genetic proclivities of students to smoke; specifically, those with higher genetic propensity to smoke are least likely to form friendships around smoking behavior in high punishment contexts. Finally, I find that smoking punishment matters not just for near-term relational outcomes, but also for downstream ones. Specifically,

individuals embedded in smoking groups and in high punishment contexts, are much more likely to smoke a decade and a half later.

Adolescent smoking and School punishment

Adolescence is a formative life stage for the development of health behaviors. According to the U.S. Department of Health and Human Services, ninety percent of adults who are regular smokers try their first cigarette before the age of 18 (USDHHS 2012). Experimenting with substances can lead to lifelong habits, influencing adult health outcomes, like lung cancer and hypertension (Boffetta & Hashibe 2006; Hecht 1999). Some students transition to harder drugs, affecting their health, academic success, addiction rates, and their likelihood of incarceration (Brook et al 1999; Annis & Watson 1975; Heitzeg 2009). These negative health behaviors are caused by, and compound with, structural inequalities of race and class, further exacerbating health disparities (Kakade et al. 2012).

In the United States, about a third of high school students reported using a tobacco product at least once in the past 30 days in 2019.¹ This number is remarkably similar to the statistics in the 1990s when 36.2% of high school students reported having used a tobacco product in the past month² (with the difference that the most common tobacco products in 2019 were e-cigarettes, compared to cigarettes in the 1990s).

With concerns of health and safety in mind, many schools punish cigarette smokers by suspending them, giving them detention, or calling their parents or guardians. By some accounts, harsh punishment can deter students from substance use. Research suggests that harsher punishment of tobacco use in schools can lead to a small, but non-trivial, reduction in smoking

¹ <https://www.cdc.gov/media/releases/2019/1205-nyts-2019.html>

² <https://www.cdc.gov/mmwr/preview/mmwrhtml/00015104.htm>

behavior on average (Botello-Harbaum et al 2009; Adams et al 2009). A long literature in sociology and criminology has established the direct deterrent effect of punishment on adolescent behaviors more generally (Pentz et al 1989), and its association with the sense of safety in that school (Farmer 1999).

Recently, scholars have begun to examine the effects of harsh punishment on the students who are punished, finding that punishment alters their life trajectories. The negative effects of punishment on those who are punished are a) many and b) concentrated among specific racial and socioeconomic groups. Boys, black students and lower socioeconomic status individuals are more at risk of being suspended or expelled from school (e.g., Welch & Payne 2010; Mizel et al. 2016 Eitle & Eitle, 2004; Anyon et al 2014; Skiba et al 2000) for reasons attributable, at least in part, to the unequal perceptions of staff and teachers (e.g. Bowditch, 1993; Downey & Pribesh, 2004). In addition to the higher rates at which male, black, and poor students experience exclusionary punitive measures, they are more likely to experience negative downstream consequences because of it (such as involvement with the criminal justice system) compared to their female, white and higher SES counterparts (Cuellar and Markowitz 2015; Wolf and Kupchick 2016). In 2013-2014 school year, 5.3% of high school students had received an out-of-school suspension³, again similar to the rate in 1992, which was 6% (Mansfield & Farris 1992).

The Relational Externalities of Punishment

Work in sociology of culture, sociology of education, and social network literatures have highlighted the role that regulatory processes, whether advanced by institutions or communities, shape network topology. The sociology of education has long established the impact of school

³ https://nces.ed.gov/programs/raceindicators/indicator_RDA.asp

tracking on adolescent social structure (Hallinan & Sorensen 1985; Lucas 2002; Gamaron 1992; Oakes 1987; Sørensen 1970). Even in the absence of formal tracking, Frank et al. (2013) show that course taking patterns provide students with local positions within which they are nearly twice as likely to form friendships. Zeng and Xie (2008) show how structural constraints in the form of grade-levels, shapes the opportunity structure of adolescent friendship formations. Similarly, Kossinets & Watts (2009) show that structural factors such as courses and email listservs, in addition to individual choice, play a crucial role in organizing the emerging network.

Generalizing the relationship between school characteristics and adolescent social networks in an era where formal tracking is not a common practice in the US, McFarland et al. (2014) show that the institutional characteristics of school, such as classroom size, electiveness versus selectiveness of courses and activities, and achievement homogeneity of classrooms moderate tie-formation micro-processes, ultimately shaping macro network structures. Social networks, as they show, cannot be explained purely as emergent outcomes of bottom-up processes such as homophily, triadic closure, and dominance without considering contextual and cultural factors. Put differently, sociologists know much more about the dynamics of diffusion given network topologies (Centola, 2018) than the social, cultural and institutional processes of real-world network formation. Substantively, McFarland et al. (2014) find that settings that could be characterized as “open and free” tend to amplify naturally present bases of group-ness such as race and age. Settings that were more structured and defined by social control, on the other hand, dampen natural modes of association and instead bring institutionally defined boundaries to the foreground of network segregation. Similarly, Frank et al. (2008) find that there is more stratification of math achievement in schools where course taking is less strictly structured.

Punishment differs from the aforementioned mechanisms of institutional control in a

number of important ways. Punishment demarcates right choices from wrong ones through the infliction of penalties for certain behaviors. Penalties—which can range from verbal warnings to expulsion in the case of smoking— are designed to harm students in specific ways that either convince them to change their behavior going forward or else serve as a warning to other students about the consequences of certain actions. Compared to other forms of control, punishment demands visibility: students should know who is punished so that they can learn the intended consequences. Further, its harshness can have particular effects on students, their perception of each other’s status and behaviors, and their tendency to conceal or not to conceal their behaviors.

The broad application of punishment in schools and the nature of its effects on student behavior may have important implications for student’s social relations that have yet to be fully explored by sociologists. For example, we may expect that through symbolic, resource-based, and locational mechanisms, punishment will lead to bifurcations in student social networks as those who engage in the punished behavior and those who do not are drawn further apart in network space. With respect to symbolic mechanisms, work in sociology of culture and criminology examines the pathways that link institutional and communal regulatory practices to identity formation and, by extension, to the formation of social relations. Labeling theory in criminology, for example, argues that punishment symbolically impacts how punished individuals perceive themselves and are perceived by others, leading to feedback loops that further separate the punished from others (Lemert 1967; Paternoster & Iovanni 1989; Bernburg & Krohn 2003). More recently, in the sociology of culture, Frye (2017) shows that for Malawian girls there are two competing pathways to adulthood: one of them based on prioritizing education and deferring marriage, and another based on prioritizing marriage and starting a family. These competing pathways, enforced by community members such as teachers and parents, lead young girls who

are thought to gravitate towards one path (say, involvement in romantic relationships), are treated as already belonging to it and are hence excluded from the other possibility of adulthood (education). Frye's work shows a complex picture of how cultural expectations and regulatory practices feed into each other to create the reality of the housewife-versus-educated-woman dichotomy that Malawian girls have to navigate. Symbolic mechanisms of this sort may lead students who smoke to become self-identified smokers, and, through school-enforced mechanisms, their own self-conception, and the perception of other students, to increasingly associate with other smokers and not with non-smokers.

Even without symbolic means, students who smoke may still be drawn into association. For example, like tracking, punishment may effectively sort students into well and ill-behaving groups by placing them in specific spaces (most commonly, detention halls). Students who smoke will then be more likely to meet as they come to occupy similar spaces in the school environment. Punishment might also limit spaces in which students can safely smoke thereby leading to higher rates of interaction and ultimately friendship. Alternatively, as when black markets for illicit substances form in the face of governmental punishment, students who smoke may come to rely on each other in the face of the harsh punishment reality for emotional support or for resources and information to support their smoking habit.

That said, the relationship between punishment and network structure with respect to smoking may go in the other direction. For example, a large literature in sociology shows how processes of stigmatization can lead to the social isolation of stigmatized individuals (Kreager 2004; Strauss & Pollack 2003). Relatedly, to avoid detection, smokers may conceal their smoking status to other students, reducing the likelihood that social relations form around smoking in a school setting.

Genes, Sociodemographic Characteristics, and Local Positions

The relationship between punishment and individual outcomes might not be the same for all students. Local positions in the network, individual attributes in the form of sociodemographic characteristics, and genetic tendencies may all moderate punishment's effects. If punishment produces inequalities in how a behavior is distributed across a social network, then individual exposure to smoking behavior may vary greatly, leading to divergences in outcomes depending on one's position in the network with regards to smoking status⁴.

Previous research has found that boys, whites and lower socioeconomic status adolescents have higher rates of smoking (National Center for Chronic Disease Prevention 1998). With regards to responsiveness to cigarette price policies, controlling for potentially confounding factors such as SES, Hispanics and blacks are more likely to quit smoking in the face of increased cigarette costs (CDC 1998). A separate literature on ethnic homophily in schools, finds that race is a more salient factor in friendship formation among Hispanic and black students (Moody 2001). Higher racial homophily and lower smoking rates among non-whites also means that there is a lower chance of diffusion of smoking behavior among non-whites. Taken together, we can expect a higher baseline smoking homophily among white students.

Smoking homophily as a result of punishment might also be mediated by gender. Previous research finds that boys tend to smoke more than girls, in part because, in many settings, smoking is less acceptable for girls than boys (Westmaas & Ferrence 2002). This higher

⁴ For example, say that harsher punishment results in adolescent groups forming around smoking behavior, such that groups of smokers and non-smokers form. In that case, knowing where an individual falls in the network – are they part of the smoker group or the non-smoker group – is important for predicting the effect punishment will have. There may also be endogenous factors that, given some initial degree of homophily, can exacerbate inequalities. DiMaggio and Garip (2011) show how network externalities (i.e. the benefit one receives from adopting a behavior depending on the number of other people in one's network who have already adopted it), over time leads to unequal outcomes on a macro-level.

degree of stigmatization of smoking for girls may mean that they need more social support for their habit to persist, especially under the threat of harsh punishment. Smoking and punishment may also have greater relational consequences for girls, such that they are more likely to lose their non-smoker friends under harsh punitive conditions. In either case, we would expect to see higher homophily in the face of harsh punishment for girls than boys. On the other hand, since boys are more likely to be punished than girls (Ahern et al 2000), they may be more reactive to the punishment environment, changing their friendship networks to be more congruent with their smoking status.

Genes may similarly mediate an individual's response to smoking punishment. Smoking has a moderate to high genetic component as measured by the proportion of its variance attributable to genetics (Li et al 2003; Boardman et al 2010). Research on the genetics of tobacco use has found that genes play an especially significant role in the number of cigarettes smoked per day and whether a person smokes regularly (Liu et al 2019; Belskey et al 2013). Studies have found that smoking has become more heritable over birth cohorts (Boardman 2010; Domingue et al 2016). The increase in the genetic component of smoking is said to be linked to the increased availability of information about smoking's harmfulness, as well as stricter restrictions placed on cigarette smoking in public spaces. Increasing the associated cost of smoking is said to have changed the composition of smokers, leading those who were less biochemically or behaviorally drawn to smoking to abandon the habit, and leaving individuals with a higher genetic tendency still smoking. Boardman et al (2010), for example, show that after the 1964 Surgeon General's report, smoking became more heritable due to the changing composition of smokers. Previous research on excise tobacco tax has also found that heavy smokers are least likely to change their smoking behavior in response to increased costs (Nesson 2017) and that this inelasticity may

have a genetic component (Fletcher 2012). Fletcher (2012), for example, finds that individuals with a lower genetic propensity to become dependent on nicotine reduced their smoking habits in response to state tobacco tax policies, while the smoking behavior of those with higher genetic propensity towards nicotine dependence was inelastic to this institutionally mandated cost.

How the genetic tendency to smoke interacts with punishment to impact friendship formation around smoking is less clear. On the one hand, we might expect high PGS individuals to smoke regardless of their friend group composition. They may be less likely to seek other smokers as friends in order to sustain their smoking habit. The relationship between their smoking behavior and friendship formation may therefore be less reactive to school punishment. On the other hand, heightened stigma may lead to the isolation of high PGS individuals (who might keep smoking in the face of increased costs) and lead to the formation of friend groups with other high PGS individuals.

Data

The data for this study comes from the National Longitudinal Survey of Adolescent to Adult Health (Add Health). Add Health is a nationally representative, longitudinal study of the health statuses and behaviors of middle and high school students in the United States collected in five waves between 1994 and 2018. During the first wave, over 90,000 students were surveyed about their health, social backgrounds, academics, friendships, and romantic relationships, among other topics. The in-school data is supplemented by a questionnaire administered to school administrators, providing overview of institutional policies and characteristics.

A subsample of roughly 20,000 students were interviewed in their homes in greater depth about their academics, health, social networks and home environments, in addition to being

given the in-school survey. Parents and other family members were also interviewed as part of the initial set of in-home surveys. Add Health has since followed the members of the in-home sample over four subsequent waves adding information about their health and social outcomes and changing life circumstances.

In Wave IV, genetic data were collected for a subset of the respondents who took part in the in-home survey (roughly 10,000 individuals). Two arrays were used for the genotyping: Illumina Human Omni1-Quad BeadChip for most respondents and the Illumina Human Omni-2.5 Quad BeadChip for the rest. The large number of SNPs tagged by these arrays – more than 1.1 million and 2.5 million, respectively – enable me to construct summary measures over the whole genome to capture students' polygenic risk of engaging in smoking behavior. To prepare the genetic data for analysis, I subjected them to a series of standard quality control procedures, filtering out SNPs with high missingness (> 0.1), low minor allele frequency (< 0.05), or which failed a Hardy-Weinberg equilibrium test ($p\text{-value} < 1e-6$).

This study draws on data from across Add Health's multiple waves to make its case. To understand network structure, it uses data on students' friendship relations, collected in the in-school survey. Students were asked to nominate up to five of their closest male and five of their closest female friends by selecting them from a list of registered students at their school. Since most nominated students in a school took the Add Health in-school survey as well, the friendship nomination data can be used to construct measures for the behaviors and attributes of a student's friends. I use these measures to estimate homophily in multilevel regression analyses.

I focus specifically on smoking, how the smoking behavior of students relates to that of their friends, and whether this relationship – i.e., homophily – is moderated by how the school punishes smoking. The main smoking variable comes from a question which asks, “During the

past twelve months, how often did you smoke cigarettes?”, with answers ranging from never to nearly every day. Since I am focused on smoking identity, I use this question to construct a variable – Daily Smoker – which measures whether a student smoked at least one cigarette every day for the past thirty days. I use a similar question from the fourth wave of the data when modeling downstream outcomes. For the measure of the smoking behavior of friends, I evaluated, for each respondent, the proportion of their friends who are daily smokers. The correlation (the Point-Biserial Correlation Coefficient) between an individual’s Daily Smoker value and that of the average of their friends is a raw measure of the overall amount of smoking homophily in the data and is fairly high at 0.38.

I leverage heterogeneity between schools in how harshly they punish first smoking offenses to understand how the harshness of punishment shapes social network relations around smoking. The measure of school punishment comes from the school administrator survey asked in Wave I. School administrators were asked about the policies and attributes of the school. They reported how they punished students for a number of offenses, and specifically for this paper, how they punished a student who was caught smoking on school grounds for the first time. Administrators reported punishments ranging from a verbal warning to expulsion. For analyses reported in the main text, I use this ordinal variable as the main measure of harshness of punishment, but I also found that the results hold when it is binarized to distinguish schools that levy harsh punishments, such as suspensions (in or out of school) or expulsions, from those that levy less harsh punishments (such as verbal warnings or minor actions such as detention) for first smoking offenses. I focus on the first offense because there is less variation in how schools punish second offenses, being uniformly harsher on repeat offenders.

Finally, for the genetic analyses, the QCed, genetic data set from Wave 4 and described

above was used. Since genes remain constant over the life course, they can be analyzed with the original data gathered during the first wave to provide insight into the relationship between gene, environment, and outcomes of interest; here, the relationship between school punishment and genetic homophily on smoking.

A number of controls are included to account for unobservable differences that may exist across schools with varying levels of smoking punishment. These factors include familial environment (married household, single parent household, maternal education, family income), familial factors that directly contribute to smoking (having an older sibling, having a household member who smokes), an individual's own age, racial and nativity status, educational factors that contribute to smoking (Grade Point Average [GPA] and Peabody Vocabulary Score [PVT]), and network degree centrality (the total number of in-degree and out-degree friends the individual has). In addition, I construct a measure of school-level punishment tendency by averaging over punishment severity for all other punished behaviors excluding smoking. In Appendix B, I show that the results also hold when less control variables are included.

Methods

I model the relationship between individuals' smoking behaviors and that of their friends using a multilevel, random effects approach. Specifically, I model the average smoking behavior of one's friends as a function of one's own smoking behavior, controlling for a number of sociodemographic and contextual variables and school random effects. The simplest homophily model would predict the behavior of one's friends as a function of one's own behavior. The estimate would conceptually be the same as assortativity, a well-known, network-based homophily measure, in which the attributes of egos are in effect correlated with those of their

alters. However, assortativity is correlational and therefore potentially confounded by other factors, at either the school or individual level, that influence the extent to which individuals are similar to one another with respect to a behavior but which do not have to do directly with that behavior itself. To address this, I use a multilevel modeling approach in order to control for confounders at multiple levels, including individual sociodemographic factors and school-level covariates, as well as to test for contextual effects, i.e. interactions between variables at the individual and school levels.

In the multilevel design, individuals are nested in schools which is modeled using random effects at the school level⁵. I include a number of individual- (such as age, race, gender, number of friends, and socioeconomic status) and school-level (such as size of the school, where it is located, and its sociodemographic composition) controls. School level variables sometimes include averages of the student body's attributes; these averages reflect school-level (and not grade-in-school-level) measures. The primary dependent variable is the average smoking behavior of one's friends – the proportion of friends that smoke every day. In modeling homophily, the main independent variable of interest is the individual's own smoking behavior. The greater its coefficient the more one's own behavior predicts that of one's friends, signaling greater homophily on that behavior. However, I am not only interested in the absolute level of homophily across schools, but the extent to which smoking homophily is moderated by the harshness with which schools punish smoking. I capture this contextual effect by interacting the school-level punishment for smoking with the individual-level variable of one's smoking behavior. The resulting interaction effect estimates the extent to which the relationship between an individual and their friends' smoking behaviors (i.e. homophily) differs depending on the

⁵ In the appendix, I report an alternative operationalization in which individuals are nested in dyads, nested in schools. Results are substantively similar and significant.

harshness of the school's smoking punishment regime. In the subsequent set of models, the interaction between individual smoking behavior and school punishment environment is interacted with a third variable, the smoking polygenic score of the individual. The coefficient of this three-way interaction can be interpreted as the extent to which the individual's receptivity to the school punishment environment is moderated by their genes.

Polygenic scores (PGS) summarize an individual's risk for an outcome given their genes. They are constructed using genome-wide association studies (GWAS), where the extent to which each SNP in the genome is associated with an outcome is evaluated using regressions on large, often population-level, data sets. The weights from those regressions capture the association between the reference allele at each SNP and the outcome. Scores for individuals can then be constructed by multiplying the number of reference alleles an individual has at each SNP location by the weight of that SNP as determined by the GWAS and then summing (or averaging) the results across all SNPs. Individuals who have multiple reference alleles (which can be a maximum of 2) at positively weighted locations will be given higher scores than individuals who have no reference alleles at such locations or multiple reference alleles at negatively weighted locations. The resulting scores distinguish individuals with high genetic propensity towards an outcome from those with low genetic propensity.

To construct smoking PGS for Add Health, I used a GWAS from the Tobacco and Genetics consortium (TAG), which runs studies associating genes with tobacco use. Along with the other variables in the study, the resulting PGS was standardized by subtracting the mean and dividing by the standard deviation. Because genes covary with ancestry and ancestry with social and cultural factors that might also contribute to smoking behavior, all models with smoking PGS following the convention in the sociogenomics literature by controlling for the first four

principal components of the genetic data⁶. Principal components capture the main axes of variation in data, which, for genetic data, have been shown in most cases to be indicative of shared ancestry (Novembre et al. 2008).

For downstream outcomes, I use a structurally similar model, i.e. a multilevel model with school random effects. The key difference is that the dependent variable of interest is the intensity of an individual's smoking behavior at Wave 4 – the number of cigarettes per month – rather than friends' smoking – proportion of friends who are daily smokers at Wave 1. This enables me to continue to account for sources of confounding that stem from differences between schools and individuals, while unifying the regression design across analyses. I use the lmer package in R, which fits multilevel models using a traditional, restricted maximum likelihood approach.

Results

I begin by descriptively exploring the association of punishment with the levels of smoking behavior and suspension in high schools. On the one hand, we may expect harsh punishment to deter smoking, leading to lower levels of smoking in the harshest punishment environments; on the other, smoking punishment harshness may be a reaction to levels of smoking, such that we observe the harshest degree of punishment where smoking levels are highest. Table 1 shows the distribution of smoking frequencies across punishment environments while Table 2 shows the average number of cigarettes smoked per day in schools with a given punishment severity.

Distributions of smoking frequency are fairly stable across levels of punishment severity.

⁶ In the canonical example, chopstick use can be predicted using genes, since people of Chinese ancestry have distinct genetic signatures while also using chopsticks to eat, even though groups differ in chopstick use for purely cultural reasons.

This is confirmed by an insignificant χ^2 test, presented at the bottom of Table 1, indicating that punishment harshness and smoking frequency are seemingly independent.

Table 1: *How often in the past year did you smoke? (Var: S59A; Proportions)*

Punishment Harshness	Never	Once or twice	Once a month or less	2 or 3 days a month	1 or 2 times a week	3 to 5 days a week	Daily	N
Verbal Warning	0.61	0.13	0.03	0.03	0.04	0.04	0.13	1024
Minor Action	0.63	0.13	0.03	0.03	0.05	0.03	0.10	2063
In-school Suspension	0.63	0.14	0.03	0.03	0.03	0.03	0.11	6156
Out-of-school Suspension	0.63	0.13	0.03	0.03	0.04	0.03	0.11	6561

$\chi^2 = 18.2$, $df = 18$, $p\text{-value} = 0.44$

Table 2 shows the number of cigarettes daily smokers report smoking per day: those in the harshest punitive environments smoke slightly more cigarettes. That said, the χ^2 for Table 1 and an ANOVA for Table 2, presented below the table, show that any observed patterns are statistically insignificant. While I do not observe any measurable difference in smoking behavior across levels of punishment, I cannot rule out the possibility that punishment had a leveling-out effect on smoking behavior, leading to the seemingly equal levels of smoking across punishment levels.

Table 2: How many cigarettes do you smoke per day? (Var: H1TO7; Means and SDs)

Punishment Harshness	Students who smoke every day			
	Mean	SD	Number of schools	Number of students
Verbal Warning	6.85	7.72	6	1024
Minor Action	6.57	8.23	16	2063
In-school Suspension	7.00	7.82	37	6156
Out-of-school Suspension	7.15	8.53	28	6561

Analysis of variance: F-value = 0.7312; p-value = 0.5332

With respect to suspension, Table 3 shows the results of a logistic regression, predicting whether students had ever been suspended in high school as a function of their smoking behaviors and the harshness with which smoking is punished in their schools. In general, smoking frequency predicts one's likelihood of being suspended. Higher smoking punishment on its own does not increase the likelihood of being suspended for all students. Rather, as the interaction in the second model shows, the effect of smoking punishment harshness on student suspension outcomes depends on the student's frequency of smoking. Specifically, as the positive coefficient signals, the more students smoke in harsher smoking punishment environments, the more likely they are to be suspended. Harsher punishment of smoking is associated with actual differences in the degree to which students are punished.

Table 3: Individual-level and Contextual Predictors of Suspension

	Suspended in High School	
	Multilevel Model	Contextual Effects Model
Smoking Frequency x Smoking Punishment (First Offense)		0.031* (0.015)
Smoking Frequency	0.238*** (0.016)	0.236*** (0.016)
Smoking Punishment Harshness (First Offense)	0.076	0.010

	(0.072)	(0.078)
Age	0.412***	0.411***
	(0.045)	(0.045)
Female	-0.371***	-0.371***
	(0.037)	(0.037)
Black	0.309***	0.305***
	(0.048)	(0.048)
Hispanic	0.077	0.077
	(0.050)	(0.050)
GPA (W1)	-0.488***	-0.488***
	(0.041)	(0.041)
PVT (W1)	-0.125**	-0.124**
	(0.045)	(0.045)
Married Household (W1)	-0.072	-0.074
	(0.047)	(0.047)
Maternal Education (W1)	-0.089*	-0.089*
	(0.042)	(0.042)
US Native	0.109*	0.110*
	(0.048)	(0.048)
Single Parent Household (W1)	0.047	0.045
	(0.044)	(0.044)
Family Income (W1)	-0.072†	-0.069†
	(0.040)	(0.040)
Household Smoker (W1)	0.160***	0.161***
	(0.034)	(0.034)
Network Centrality (W1)	-0.124**	-0.126**
	(0.040)	(0.040)
Has Older Siblings (W1)	-0.035	-0.036
	(0.038)	(0.038)
Mean Age (W1)	0.395***	0.388***
	(0.103)	(0.103)
Proportion Female (W1)	0.100	0.102
	(0.070)	(0.069)
Proportion Black (W1)	-0.023	-0.019
	(0.172)	(0.172)
Proportion Hispanic (W1)	-0.253	-0.255
	(0.193)	(0.193)
Mean GPA (W1)	-0.242*	-0.239*
	(0.098)	(0.098)
Mean PVT (W1)	0.069	0.079
	(0.176)	(0.176)
Proportion Married Household (W1)	-0.004	0.001
	(0.198)	(0.198)
Mean Maternal Education (W1)	-0.340*	-0.340*
	(0.158)	(0.158)
Mean Family Income (W1)	0.195	0.191
	(0.144)	(0.143)
Proportion w. Household Smoker (W1)	-0.181	-0.187
	(0.135)	(0.135)
Mean Degree (W1)	-0.093	-0.094
	(0.097)	(0.096)
Proportion w. Older Siblings (W1)	-0.016	-0.012
	(0.097)	(0.097)
Proportion Daily Smoker (W1)	0.049	0.058

	(0.122)	(0.121)
Proportion US Native (W1)	0.047	0.044
	(0.136)	(0.135)
Proportion in Single Parent Household (W1)	0.138	0.144
	(0.182)	(0.182)
Punishment Profile (W1)	0.091	0.087
	(0.071)	(0.071)
Number of Classmates in In-home Sample (W1)	-0.095	-0.085
	(0.134)	(0.133)
School Size (W1)	-0.166	-0.169
	(0.107)	(0.106)
Region: Midwest (W1)	-0.056	-0.069
	(0.239)	(0.238)
Region: South (W1)	-0.385	-0.380
	(0.252)	(0.251)
Region: Northeast (W1)	-0.112	-0.118
	(0.249)	(0.248)
Constant	-1.988***	-1.971***
	(0.322)	(0.321)
Observations	8,486	8,486
LL	-2,671.489	-2,669.352
AIC	5,420.979	5,418.703
BIC	5,695.779	5,700.550
ANOVA over multilevel model, p-value	-	0.039*

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

The relational consequences of smoking punishment

In the next set of analyses, I explore relational consequences of smoking punishment.

Specifically, and as discussed in the previous sections, I explore the extent to which homophily on smoking status varies across schools with the expectation that homophily will be most pronounced in harsh punishment environments.

To formally test this expectation, I run a series of multilevel models assessing whether homophily varies across schools with different smoking punishments. I model homophily as when an individual's smoking frequency predicts that of their friends, and specifically, the proportion of their friends who also smoke regularly (though see Appendix A for school-level models as robustness checks). Following the contextual effects literature, I use multilevel modeling with school random effects and an interaction between a context and individual level variable (ego-centric smoking homophily and school punishment). The results can be found in

Table 4. In addition to behavioral homophily, I also present results where I assess homophily with respect to smoking polygenic score. Given that the primary treatment variable is a school's policy towards punishing smoking behavior, in Appendix E, I present balancing tests, which find that schools with different levels of punishment harshness have similar demographic compositions and state-level excise taxes on average.

The first regression (Model 1) only includes individual-level controls such as ego's race, gender, academic achievement, network centrality (measured as the sum of the number of students they nominated and the number of students who nominated them as friends), and socioeconomic status, in addition to school-level random effects and the main independent variable of interest: whether the ego is a daily smoker (defined as smoking at least one cigarette every day for the past 30 days). Model 2 (the multilevel model) adds school covariates, including the average of students' sociodemographic characteristics, average number of regular smokers in the school, school size, school's smoking punishment, average network degree, and school's punishment profile (a summary measure of a school's tendency to punish other behaviors)⁷, in addition to individual-level ones. Model 3 is the contextual effects model, capturing the interaction between individual-level and school-level characteristics, namely, whether harshness of school punishment leads to homophily on smoking behavior. Thus, the third model tests my main hypotheses, while the first and second models are useful for comparison and understanding the behavior of the coefficients as model complexity increases. Models 4 and 5 add smoking PGS to the interaction between daily smoker and smoking punishment as well as genetic controls. They differ only in their sample: Model 4 includes all respondents with genetic

⁷ This measure is constructed by summing the punishment severity for the first offense of the following behaviors as reported by the school administrator in wave 1 of Add Health: cheating or plagiarism, drinking, doing drugs, bringing weapons to school, verbally abusing teachers or other students, fighting, and stealing. The measure excludes smoking punishment.

information while model 5 includes only non-Hispanic whites. I conduct separate analyses for whites because they comprise the largest, genetically-homogenous subpopulation in Add Health. Further, the GWAS used to construct the polygenic score for smoking was fit using a sample of individuals with European ancestry. Predictive validity of the PGS is therefore expected to be highest for whites. That said, using only whites leads to a substantial decrease in the number of respondents included in the regression, which may lead the white-only analyses to be underpowered.

Table 4: The effect of school punishment on homophily and the role of genes

			Proportion of Friends who Smoke			
	Individual- level	School-level	Multilevel	Contextual Effects	Contextual Effects x PGS	Contextual Effects x PGS (Whites)
Daily Smoker x Smoking Punishment (First Offense)				0.021* (0.010)	0.050** (0.015)	0.057* (0.029)
Daily Smoker x Smoking Punishment (First Offense) x Smoking PGS					-0.039* (0.017)	-0.048 (0.034)
Daily Smoker x Smoking PGS					0.081*** (0.016)	0.046 (0.034)
Smoking Punishment (First Offense) x Smoking PGS					-0.005 (0.015)	-0.005 (0.036)
Daily Smoker	0.298*** (0.010)		0.294*** (0.010)	0.294*** (0.010)	0.267*** (0.016)	0.275*** (0.029)
Age	0.102*** (0.010)		0.100*** (0.011)	0.100*** (0.011)	0.110*** (0.015)	0.171*** (0.021)
Female	0.024** (0.009)		0.025** (0.009)	0.025** (0.009)	0.030* (0.013)	0.049** (0.018)
Black	-0.154*** (0.012)		-0.153*** (0.013)	-0.153*** (0.013)	-0.009 (0.052)	
Hispanic	-0.038** (0.013)		-0.023† (0.013)	-0.023† (0.013)	0.033 (0.027)	
GPA	-0.091*** (0.011)		-0.091*** (0.011)	-0.090*** (0.011)	-0.090*** (0.015)	-0.100*** (0.021)
Peabody Vocab. Test	-0.022† (0.012)		-0.024* (0.012)	-0.025* (0.012)	-0.022 (0.017)	-0.073** (0.026)
Parents are Married	-0.036** (0.013)		-0.035** (0.013)	-0.036** (0.013)	-0.027 (0.019)	-0.047 (0.029)
Single Parent Household	0.009 (0.013)		0.008 (0.013)	0.008 (0.013)	0.016 (0.018)	0.021 (0.028)
Maternal Education	-0.010 (0.011)		-0.007 (0.011)	-0.007 (0.011)	-0.009 (0.015)	-0.015 (0.022)

US Native	0.027*	0.023*	0.023*	0.045*	0.038
	(0.011)	(0.011)	(0.011)	(0.020)	(0.052)
Family Income	0.002	0.003	0.004	-0.003	-0.001
	(0.010)	(0.010)	(0.010)	(0.015)	(0.021)
Smoker in household	0.065***	0.060***	0.060***	0.057***	0.067***
	(0.009)	(0.009)	(0.009)	(0.013)	(0.017)
Network Centrality	-0.035***	-0.040***	-0.040***	-0.059***	-0.068***
	(0.010)	(0.010)	(0.010)	(0.014)	(0.019)
Has Older Siblings	0.0004	0.001	0.001	0.006	0.012
	(0.010)	(0.010)	(0.010)	(0.015)	(0.022)
All PC 1				0.146**	
				(0.056)	
All PC 2				-0.009	
				(0.023)	
All PC 3				0.046*	
				(0.023)	
All PC 4				0.008	
				(0.016)	
White PC 1					0.269*
					(0.114)
White PC 2					0.029
					(0.080)
White PC 3					-0.001
					(0.025)
White PC 4					0.025
					(0.075)
Smoking Punishment (First Offense)	-0.005	-0.011	-0.011	0.014	0.003
	(0.016)	(0.016)	(0.016)	(0.020)	(0.035)
Smoking PGS				0.033	0.025
				(0.028)	(0.039)
<i>School-level Variables</i>					
Mean Age	-0.009	-0.058**	-0.059**	-0.079***	-0.059†
	(0.018)	(0.019)	(0.019)	(0.024)	(0.033)
Proportion Female	0.014	0.022†	0.023†	0.022	-0.005
	(0.014)	(0.013)	(0.013)	(0.017)	(0.021)
Proportion Black	0.010	0.089*	0.089*	-0.008	-0.051
	(0.040)	(0.039)	(0.039)	(0.152)	(0.245)
Proportion Hispanic	-0.050	-0.044	-0.045	-0.081	-0.091
	(0.042)	(0.041)	(0.041)	(0.061)	(0.089)
Mean GPA	0.024	0.041*	0.041*	0.023	0.003
	(0.020)	(0.020)	(0.020)	(0.024)	(0.034)
Mean Peabody Vocab. Test	0.027	0.012	0.012	0.053	0.037
	(0.037)	(0.036)	(0.036)	(0.053)	(0.084)
Proportion Parents are Married	-0.031	-0.041	-0.040	-0.034	-0.076
	(0.045)	(0.043)	(0.043)	(0.053)	(0.071)
Mean Maternal Education	-0.027	-0.009	-0.009	-0.016	-0.020
	(0.034)	(0.033)	(0.033)	(0.041)	(0.064)
Mean Family Income	0.001	0.0002	-0.001	-0.013	0.005
	(0.029)	(0.027)	(0.027)	(0.031)	(0.044)
Proportion w. Household Smoker	0.013	-0.005	-0.006	-0.045	-0.015
	(0.030)	(0.028)	(0.028)	(0.035)	(0.046)
Mean Network Centrality	0.036†	0.054**	0.054**	0.071**	0.054
	(0.021)	(0.020)	(0.020)	(0.025)	(0.038)

Proportion w. Older Siblings		0.017 (0.022)	0.015 (0.022)	0.015 (0.021)	-0.004 (0.029)	-0.014 (0.040)
Proportion Daily Smoker		0.280*** (0.034)	0.210*** (0.032)	0.211*** (0.032)	0.248*** (0.039)	0.155** (0.055)
Proportion US Native		-0.029 (0.030)	-0.034 (0.029)	-0.034 (0.029)	-0.046 (0.045)	0.070 (0.070)
Proportion in Single Parent Household		0.021 (0.042)	-0.011 (0.040)	-0.010 (0.040)	0.017 (0.048)	0.002 (0.064)
Punishment Profile		-0.003 (0.016)	-0.008 (0.015)	-0.008 (0.015)	-0.021 (0.018)	-0.027 (0.023)
Mean Smoking PGS					0.173 (0.208)	-0.084 (0.260)
Mean All PC 1					-0.272 (0.271)	
Mean All PC 2					-0.073 (0.074)	
Mean All PC 3					-0.021 (0.032)	
Mean All PC 4					0.050 (0.042)	
Mean White PC 1						-0.198 (0.387)
Mean White PC 2						0.119 (0.120)
Mean White PC 3						0.152 (0.094)
Mean White PC 4						0.021 (0.052)
Number of Classmates in In-home Sample		0.024 (0.029)	0.028 (0.027)	0.029 (0.027)	-0.009 (0.039)	0.066 (0.049)
School Size		0.017 (0.023)	0.038† (0.023)	0.039† (0.022)	0.067* (0.029)	0.048 (0.038)
Region: Midwest		-0.024 (0.052)	0.021 (0.049)	0.020 (0.049)	0.035 (0.062)	-0.013 (0.079)
Region: South		-0.026 (0.053)	0.014 (0.050)	0.015 (0.050)	0.039 (0.067)	0.020 (0.094)
Region: Northeast		-0.050 (0.055)	-0.025 (0.053)	-0.025 (0.053)	-0.003 (0.072)	0.019 (0.090)
Constant	0.041† (0.022)	-0.018 (0.072)	-0.081 (0.069)	-0.084 (0.068)	-0.163† (0.085)	-0.116 (0.139)
Observations	8,773	11,420	8,704	8,704	4,387	2,776
Log-Likelihood	-11,073.060	-15,658.470	-10,888.170	-10,885.890	-5,541.612	-3,735.03
AIC	22,182.120	31,366.950	21,856.350	21,853.790	11,191.22	7,574.07
BIC	22,309.550	31,550.530	22,139.210	22,143.720	11,536.09	7,882.36
ANOVA over multilevel model, p-value	-	-	-	0.033*	8.52E-07***	0.21

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Across all models, there is a strong relationship between an individual's smoking behavior and

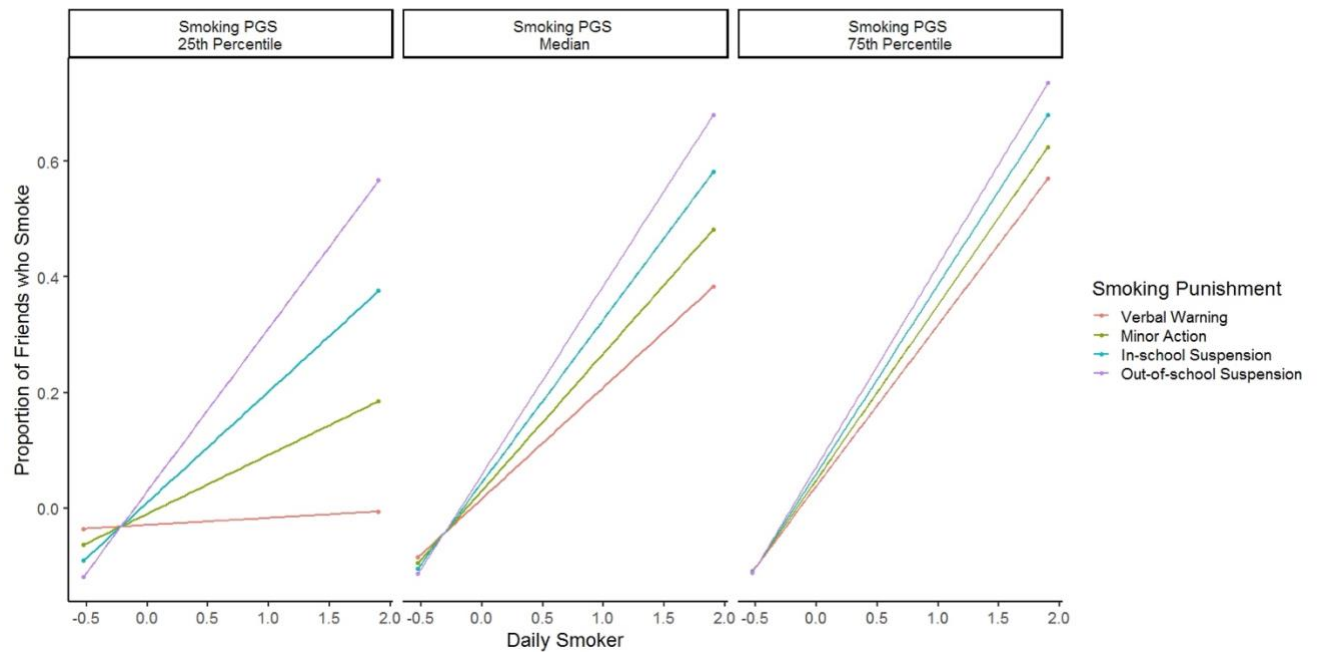
that of their friends, and this relationship is robust to the addition of controls. Models 1 and 2 reveal other individual and school-level covariates which predict the smoking behavior of a student's friends. Mean age appears to be a significant and negative predictor of the portion of one's friends who smoke daily. The heterogeneity in this variable is likely due to heterogeneity in how high schools are structured: for example, not all schools in Add Health have all four high school grades. Some schools are only made up of 11th and 12th grades, while others have 9 to 12th.

At the individual-level, being older, female, a US native, and having a household member who smokes, positively predict having friends who smoke, while being black or Hispanic, having a higher GPA, having parents who are married (compared to divorced households), or having a central position in the school network, negatively predict having friends who smoke. The consistent negative association of network centrality and friend smoking shows that smoking is fairly marginalized activity – popular students are less likely to have smoker friends. The fact that girls have a higher baseline homophily for smoking is suggestive of the theory that they may require more social support for smoking, but this will be more directly addressed in the next section.

At the school-level (Model 2), many of the same associations are found. For model 3, the independent variable of interest is the interaction between daily smoker and smoking punishment harshness, which captures the extent to which the degree of homophily is moderated by the harshness of smoking punishment in schools. The coefficient is positive, signaling that as smoking punishment increases, the strength of the relationship between an individual's smoking behavior and that of their friends increases. In other words, there is greater homophily in the harsher smoking punishment environments.

With respect to the genetic, three-way interactions, I find a significant and negative coefficient for the sample with all students, and an insignificant coefficient of similar sign and magnitude among whites. Recall that a higher polygenic score means higher genetic propensity for smoking. The negative coefficient implies that the effect of school punishment on homophily (i.e., the interaction between daily smoker and school punishment) is less pronounced for those with high smoking PGS than low smoking PGS. The three-way interaction is visualized in Figure 2, where the extent to which smoking punishment moderates the slope of the relationship between being a daily smoker and proportion of one's friends who smoke is compared at three different terciles of smoking PGS. Across all three terciles, daily smokers in harsh punishment environments have higher homophily than non-smokers. However, in the 25th percentile of smoking PGS, we see that homophily varies greatly by the punishment environment, while in the 75th percentile, differences are very modest. This aligns with the theory that the behavior of those with high genetic propensity is less elastic to policy effects, which has been found for other outcomes, including smoking behavior (though not smoking homophily). In a similar vein, it signals that low PGS individuals, whose smoking behavior may be more malleable, need more social support to become regular smokers.

Figure 2: Smoking PGS modulates the effect of smoking punishment on smoking homophily

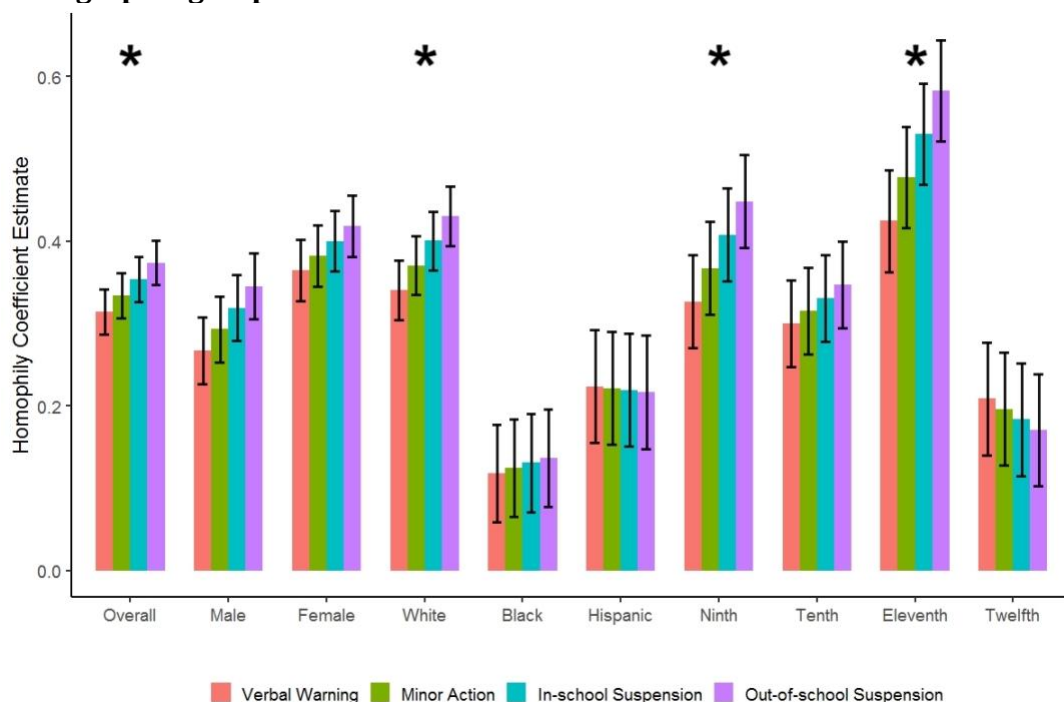


Heterogeneities in Smoking Homophily

To assess whether or not there is heterogeneity in how smoking homophily responds to punishment by demographic characteristics, such as race, gender and grade, I run the same models as above but subset the sample to include only members of a given demographic group. The results are presented in Figure 1. Here, the coefficient at each level of punishment is plotted for each group with asterisks signifying whether the interaction between daily smokers and smoking punishment is significant. For most groups, the core results, in which homophily increases in harsher punishment environments, hold. Results are significant overall, among whites, and among ninth and eleventh graders. Three groups are notable exceptions. Hispanic and black students both show low levels of overall homophily on smoking with little to no difference under lenient or strict punishment environments. Black and Hispanic students are less likely than white students to smoke, which may limit the potential for homophily on smoking. Further, their identities as Hispanic and black students, especially in racially diverse or majority

white schools, may eclipse other characteristics in influencing friendship formation.

Figure 1: How homophily and the effect of the punishment environment differ by demographic group



Twelfth graders have similarly low homophily (but note that the interaction between daily smoker and punishment among twelfth graders is insignificant, indicating that the negative slope shown in Figure 1 is not significantly different from zero). Of the four high school grades, only twelfth graders can legally smoke tobacco. This change in the legality of smoking behavior may change a) the extent to which smoking behavior is treated as transgressive, b) the need for smokers to share resources and offer support to one another in the face of punishment, and c) the likelihood that students will be punished harshly for their smoking behavior, all of which could reduce smoking homophily, as networks built around smoking identity and support in the face of punishment no longer serve a need. Tables in Appendix C show that indeed 12th graders are less likely to be suspended than their younger peers, which may signal that administrators are more lenient towards their smoking behavior given its legality, though the Add Health data do not

allow us to evaluate whether this is due to fewer suspensions for smoking or a general leniency across schools in how twelfth graders are treated.

Downstream consequences of punishment

Thus far, I have shown that school punishment on smoking impacts smoking homophily and that this is moderated by gender, grade, and smoking PGS. In this section I examine whether smoking punishment and smoking homophily influence students' downstream outcomes, and specifically, the intensity of their smoking behavior (the number of cigarettes they smoked in the past month) their as adults as measured in wave IV, roughly 15 years after the initial Wave I interviews. I also examine whether these downstream effects are at all mediated by respondents' genetic propensity to smoke. The regressions assess the contribution of friend smoking behavior, punishment, smoking PGS, and their interaction, for all respondents in an initial set of regressions and for only those who had tried smoking in Wave 1 in a second set.

The rationale for including the interaction between punishment and friend smoking behavior is as follows: since school punishment is a context-level factor, it can have different effects on students in different social circles. The interaction between friend-group smoking levels and punishment will allow us to assess how punishment specifically affects students in friend groups where smoking was prevalent. As shown in the previous sets of results, harsher punishment results in higher homophily, which we can expect to result in greater inequality in future smoking outcomes between those who smoke and those who do not. Focusing solely on the direct effect of punishment would not allow us to capture such differences.

Table 5: The effects of smoking punishment on downstream smoking behavior

	Smoking behavior (Num. of cigs. in past month [W4])			
	Without Interaction Effects	Main Interaction Model	With PGS, All Respondents	With PGS, White Respondents
Proportion of Friends who Smoke x Smoking Punishment (First Offense)		0.283* (0.142)	0.118 (0.190)	0.072 (0.375)
Proportion of Friends who Smoke x Smoking Punishment (First Offense) x Smoking PGS			0.089 (0.186)	0.221 (0.439)
Proportion of Friends who Smoke x Smoking PGS			0.282 (0.218)	0.673 (0.448)
Smoking Punishment (First Offense) x Smoking PGS			0.194 (0.202)	0.689 (0.444)
Proportion of Friends who Smoke	0.663*** (0.159)	0.641*** (0.159)	0.474* (0.215)	0.269 (0.389)
Smoking Punishment (First Offense)	-0.087 (0.223)	-0.089 (0.224)	-0.136 (0.299)	-0.372 (0.440)
Smoking PGS			1.095** (0.367)	0.795 (0.492)
Years of Education	-1.108*** (0.077)	-1.111*** (0.077)	-1.137*** (0.098)	-1.223*** (0.130)
Logged Income	-0.065 (0.052)	-0.063 (0.052)	-0.096 (0.066)	-0.038 (0.085)
Daily Smoker (W1)	3.635*** (0.159)	3.629*** (0.159)	3.814*** (0.199)	3.916*** (0.242)
Age	-1.033*** (0.160)	-1.022*** (0.160)	-1.195*** (0.202)	-1.238*** (0.260)
Female	-0.441** (0.143)	-0.441** (0.143)	-0.448* (0.181)	-0.298 (0.239)
Black	-0.097 (0.195)	-0.101 (0.195)	0.255 (0.681)	
Hispanic	-0.384† (0.198)	-0.379† (0.198)	-0.044 (0.357)	
GPA (W1)	-0.593*** (0.169)	-0.588*** (0.169)	-0.731*** (0.213)	-0.505† (0.280)
PVT (W1)	0.719***	0.720***	0.698**	0.566†

	(0.175)	(0.175)	(0.226)	(0.323)
Married Household (W1)	-0.136	-0.143	-0.337	-0.842*
	(0.201)	(0.201)	(0.253)	(0.363)
Maternal Education (W1)	0.319*	0.318*	0.235	0.374
	(0.159)	(0.159)	(0.204)	(0.282)
US Native	0.005	0.001	-0.061	-0.905
	(0.181)	(0.181)	(0.259)	(0.649)
Single Parent Household (W1)	-0.051	-0.055	-0.272	-0.681†
	(0.193)	(0.192)	(0.242)	(0.355)
Family Income (W1)	0.049	0.054	-0.164	-0.163
	(0.152)	(0.152)	(0.199)	(0.271)
Household Smoker (W1)	0.747***	0.747***	0.705***	0.870***
	(0.132)	(0.132)	(0.168)	(0.220)
Network Centrality (W1)	-0.108	-0.106	0.002	-0.325
	(0.150)	(0.150)	(0.190)	(0.239)
Has Older Siblings (W1)	-0.152	-0.160	-0.199	-0.448
	(0.153)	(0.153)	(0.196)	(0.273)
All PC 1			-0.472	
			(0.733)	
All PC 2			-0.063	
			(0.309)	
All PC 3			0.361	
			(0.301)	
All PC 4			-0.010	
			(0.205)	
White PC 1				0.513
				(1.424)
White PC 2				0.289
				(0.989)
White PC 3				0.473
				(0.312)
White PC 4				-0.996
				(0.924)
Mean Age (W1)	-0.511†	-0.537†	-0.191	-0.358
	(0.277)	(0.278)	(0.370)	(0.428)
Proportion Female (W1)	0.197	0.197	-0.082	0.034
	(0.192)	(0.192)	(0.253)	(0.269)
Proportion Black (W1)	0.404	0.402	4.788*	5.907†
	(0.559)	(0.561)	(2.368)	(3.094)
Proportion Hispanic (W1)	-0.247	-0.275	0.581	-0.598

	(0.592)	(0.595)	(0.926)	(1.139)
Mean GPA (W1)	0.199	0.192	0.566	0.559
	(0.284)	(0.285)	(0.382)	(0.438)
Mean PVT (W1)	0.480	0.514	-0.488	0.609
	(0.563)	(0.565)	(0.916)	(1.117)
Proportion Married Household (W1)	0.088	0.099	0.008	-0.157
	(0.615)	(0.617)	(0.830)	(0.908)
Mean Maternal Education (W1)	-0.751	-0.757	-0.784	-1.597†
	(0.466)	(0.468)	(0.652)	(0.819)
Mean Family Income (W1)	0.664†	0.642†	0.441	0.500
	(0.377)	(0.379)	(0.501)	(0.554)
Proportion w. Household Smoker (W1)	0.318	0.290	0.390	0.049
	(0.398)	(0.400)	(0.532)	(0.585)
Mean Degree (W1)	0.244	0.244	0.172	0.172
	(0.286)	(0.287)	(0.387)	(0.475)
Proportion w. Older Siblings (W1)	0.056	0.078	-0.333	-0.253
	(0.317)	(0.318)	(0.441)	(0.514)
Proportion Daily Smoker (W1)	0.852†	0.875†	0.636	0.422
	(0.461)	(0.463)	(0.602)	(0.696)
Proportion US Native (W1)	-0.446	-0.454	-0.135	-1.579†
	(0.427)	(0.428)	(0.697)	(0.887)
Proportion in Single Parent Household (W1)	0.422	0.458	0.038	0.530
	(0.582)	(0.584)	(0.755)	(0.828)
Punishment Profile	-0.089	-0.096	-0.050	-0.262
	(0.214)	(0.214)	(0.285)	(0.296)
Mean Smoking PGS			-1.613	-1.904
			(3.188)	(3.281)
Mean All PC 1			5.980	
			(4.168)	
Mean All PC 2			-0.503	
			(1.145)	
Mean All PC 3			0.322	
			(0.487)	
Mean All PC 4			-0.124	
			(0.708)	
Mean White PC 1				9.544†
				(4.888)
Mean White PC 1				-1.754
				(1.526)

Mean White PC 1				-2.225† (1.181)
Mean White PC 1				0.566 (0.668)
Number of Classmates in In-home Sample (W1)	0.121 (0.364)	0.136 (0.366)	-0.248 (0.696)	-0.693 (0.635)
School Size (W1)	-0.367 (0.317)	-0.357 (0.319)	-0.606 (0.458)	-0.502 (0.483)
Region: Midwest (W1)	0.155 (0.686)	0.122 (0.688)	0.706 (0.966)	1.201 (0.999)
Region: South (W1)	1.105 (0.698)	1.130 (0.701)	0.911 (1.031)	2.528* (1.194)
Region: Northeast (W1)	-0.558 (0.753)	-0.558 (0.755)	0.067 (1.128)	-0.602 (1.148)
Constant	25.630*** (1.535)	25.622*** (1.538)	27.137*** (2.039)	26.654*** (2.671)
Observations	6,495	6,495	4,226	2,716
LL	-24,679.710	-24,678.760	-16,128.040	-10,454.500
AIC	49,445.420	49,445.530	32,370.070	21,019.000
BIC	49,736.910	49,743.790	32,731.970	21,343.880

The results can be found in Table 5. I include a relatively large set of school-level and individual-level controls to avoid confounding by demographic differences between schools and individuals. In Appendix B, I show that the results (both homophily and downstream) are the same even without such controls. Unsurprisingly, the single strongest predictor of smoking in Wave 4 is whether one was a daily smoker in Wave 1. Smoking persists over the life course, a fact attributable at least in part to addiction. However, even controlling for past smoking behavior, the smoking behavior of one's friends exhibits a strong, independent influence on one's downstream smoking behavior, long after respondents share a school environment with their friends. Smoking punishment, on the other hand, does not exhibit an independent effect; rather, its effect is in its interaction with friend smoking behavior. As mentioned above, smoking

punishment may have opposite effects on the smoking behaviors of individuals embedded in different friend groups. Looking only at smoking punishment of schools may mask important heterogeneity originating from the local friend environment, while looking only at friend groups will miss the impact that the institutional context has on peer effects. The interaction between the local situation of the student (with regards to smoking), combined with the punishment regime of the school, provides a clearer picture. For those in harsh punishment environments, the smoking behavior of friends has an outsized influence on their own future behavior.

Finally, for the downstream outcomes, genes do not appear to play a role in moderating the effects of school punishment harshness or friend smoking behavior. Smoking PGS predicts downstream smoking behavior but does not significantly interact with the two main independent variables.

Discussion

This paper considers the relational consequences of institutional control, expressed in the punishment of smoking behavior in high schools. Harsher policies towards smoking behavior in schools have little to no direct association with student's smoking behavior, whether their frequency of smoking or, for smokers, the number of cigarettes they smoke per day.

Smokers cluster together in the face of the threat of punishment. Homophily on smoking is reasonably high across schools and it is highest in schools where smoking is harshly punished. How students react to the school punishment environment when forming friendships is moderated by a student's own genetic risk of smoking: those with the highest genetic propensity to smoke are least responsive to school punishment. This aligns with findings in public health that heavy smokers are least responsive to increases in the costs of cigarettes (Nesson 2017) and

that this inelasticity may have a genetic component (Fletcher 2012). Lower PGS individuals may be more susceptible to changing their smoking behavior or need a stronger social support system to sustain it. Put differently, there may be more network externalities for lower PGS individuals, who may be unlikely to start smoking in the absence of exposure in their peer network.

The effect of school punishment on homophily holds for almost all subgroups in the data with notable exceptions of black and Hispanic students and twelfth graders. How do we make sense of the exceptions? Smoking culture is more pervasive among white students than black and Hispanic students, with initiation to smoking happening earlier and more often for white students (Kandel et al. 2004). Further, black and Hispanic students are far more likely to be minorities in their high school settings than white students, a function of their minority status in the country as a whole, despite persistent segregation by race. This may lead racial identity to be a more salient basis of friendship formation than smoking among minority students, regardless of the punitive environment. From the perspective of white students, given that most of their peers are white, they will be better positioned to select friends on smoking status without having to compromise on their within-race preferences.

The marked drop in homophily between eleventh and twelfth grade may be due to a relaxation of the punishment environment - at both the school and county level. After all, by twelfth grade, the majority of students can legally use tobacco. A cursory glance at suspension numbers and smoking rates by grade, presented in Appendix C, reveals that twelfth graders are less likely than eleventh graders to be suspended. In theory, this could reflect changes in student behavior as students approaching graduation behave more cautiously to avoid jeopardizing their status. However, when it comes to smoking at least, twelfth graders are as likely, if not more likely, than eleventh graders to smoke. In part then, we may be seeing a relaxation of the punitive

environment, particularly concerning smoking, as students approach the end of high school.

The relational externalities of school punishment matter for students' outcomes later in life too. Those who were embedded in networks with more smokers were more likely to smoke as adults. But this effect was even more pronounced in highly punitive schools. Interpreted one way, the school punishment environment may increase the salience of friends' smoking behavior for one's own behavior, which through addiction, leads to downstream smoking. Interpreted the other way, the smoking punishment environment matters most for students who are friends with smokers, perhaps because they are most exposed to punishment and its effects.

As the previous paragraph alludes, the potential set of mechanisms underpinning the results are many. Students may be brought into association according to their smoking behavior in high punishment environments for a number of reasons. Identity or labelling theories might argue that harsh punishment leads smoking to become more socially salient and a marker of identity, distinguishing smokers from non-smokers. This could lead smokers to seek out other students who share their habit, and by extension identity, or non-smokers to avoid smokers, for fear of being punished or labeled as one.

Homophily may also be driven by resource sharing. In contexts where access to cigarettes is limited and riskier to obtain, students may rely on each other to obtain cigarettes clandestinely. Similarly, harsh punishment may encourage secrecy among smokers, necessitating trust and the formation of close and exclusive ties. Harsh punishment may constrain the number of spaces where smoking is possible, such that smokers may increasingly find themselves clustered in those locational foci (Feld 1981), encouraging friendship formation. Finally, smokers may seek out friends that share their behavior for solidarity with smokers providing support systems that foster further engagement with the behavior. Solidarity may be even more important for groups

of people for whom, or in contexts in which, smoking is risky or viewed as inappropriate or taboo.⁸

Add Health lacks the data to definitively adjudicate between these competing mechanisms; however, the results can still help us make sense of what may be happening. 1) The fact that by twelfth grade punishment is not associated with clustering around smoking, in addition to twelfth graders' lower rates of suspension in general, suggests that the mechanism underlying smoking homophily has more to do with actual exposure to punishment than the symbolic effects of the threat of punishment on identity formation, which we would not expect to have a gradient with respect to grade. Regressions in Appendix D lend more evidence to this interpretation by showing that the core results hold even when the number of smokers punished in a school is used as the measure of how harshly a school punishes smoking. 2) The findings regarding PGS and female students provide suggestive evidence that the operative mechanism is social support. Groups whose smoking status is more amenable to change require a social support system to smoke. The difference between twelfth grade and other grades extends this finding to contexts – in contexts where smoking is riskier (i.e., more likely to lead to punishment), students require more social support to smoke. This does not necessarily exclude other, more practical reasons for homophily – locational foci might provide local contexts where support for smoking is higher; resource sharing is itself an expression of support. This dynamic could also be due to influence – for those and in the contexts where smoking behavior is less risky, friends are more likely to influence their friends' smoking behaviors. 3) The downstream results provide an interesting temporal angle. Although I find no impact of punishment on smoking homophily in twelfth grade, punishment and homophily have an impact on smoking in

⁸ I thank Paul DiMaggio for pointing out these alternative possible reasons for clustering.

adulthood. This kind of temporal lag is present in all adolescent peer effects literature. Students will eventually leave school, but habits, friends, and identities adopted in school affect students' later life outcomes. In this case, higher exposure to cigarettes in adolescence may contribute to tobacco consumption habits and addiction, which last through adulthood. 4) The effects of school punishment are not equal for everyone, but they are mediated through the individual's local social context. There is a robust association between friends' smoking behavior and one's own downstream smoking, regardless of their smoking habit in Wave 1. But this effect is even more strong for those in highly punitive schools. While previous studies found that adolescents' friends matter for their adult outcomes (Ali & Dwyer 2009), I show that the contexts within which adolescents are embedded also matter, in addition to their friendships, and that the effect of punishment is mediated by their friend groups.

Future work can build on these results by adjudicating more precisely between the proposed mechanisms that link harsher punishment environments with network homophily. The bulk of the evidence here supports the theory that the environment produced by harsh punishment leads student smokers to seek social and practical support for their habit. However, due to a lack of data, theories which focus on locational factors, such as sharing detention or similar locations for smoking, or identity-based mechanisms could not be fully evaluated. This will require data sources rich not only in network heterogeneity across institutions, but in information about respondent's attitudes towards and motivations for their different friendships.

With such data, the exact role that genes play in the unfolding of social relations in the face of institutional control can also be better understood alongside social causes. Sociologists are just beginning to explore how genetic tendencies shape networks and tie formation and dissolution processes. Much of network research has focused on acquired and therefore

malleable attributes — hence, the great deal of research dedicated to studying influence and diffusion. Genes are interesting for network research precisely because they are unchanging and exist prior to tie formation. Further, if genes indeed influence our tastes and the likelihood of liking and becoming addicted to substances or behaviors, they can be incorporated into traditional network models, such as threshold models or models of structural balance, where people's receptivity towards new behaviors and preferences towards existing ones are important determinants of network outcomes.

Overall, this paper brings together perspectives from the fields of criminology, social networks, organizations, sociogenomics to understand the institutional conditions under which behaviors are taken up and social relations are formed around them. Institutions do more than impact an individual's probability to engage in a behavior; they shape the relational context around behaviors. Sociologists are just beginning to understand how and in what ways institutions shape networks. This paper complicates the story further by showing how the relational effects of institutions may vary systematically with differences in individuals, whether genetic or sociodemographic. As this paper demonstrates, only by considering the multiple layers of the social system, from the institution (macrolevel) to the network (mesolevel) to the individual (microlevel) can we understand behavior in action and in context.

References

- Adams, M. L., Jason, L. A., Pokorny, S., & Hunt, Y. (2009). The relationship between school policies and youth tobacco use. *Journal of School Health, 79*(1), 17-23.
- Ahern, C. H., Batchelor, S. M., Blanton, C. J., Law, M., Loo, C. M., Pevzner, E. S., ... & Robb, W. H. (2000). Youth tobacco surveillance—United States, 1998–1999. *MMWR Surveill Summ, 49*(SS-10).
- Ali, M. M., & Dwyer, D. S. (2009). Estimating peer effects in adolescent smoking behavior: a longitudinal analysis. *Journal of Adolescent Health, 45*(4), 402-408.
- Annis, H., & Watson, C. (1975). Drug use and school dropout: A longitudinal study. *Canadian Journal of Counselling and Psychotherapy, 9*(3-4).

- Anyon, Y., Jenson, J. M., Altschul, I., Farrar, J., McQueen, J., Greer, E., ... & Simmons, J. (2014). The persistent effect of race and the promise of alternatives to suspension in school discipline outcomes. *Children and Youth Services Review*, 44, 379-386.
- Ascani, N. (2012). Labeling theory and the effects of sanctioning on delinquent peer association: A new approach to sentencing juveniles. *Perspectives*, 4(1), 10.
- Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences*, 115(42), E9765-E9772.
- Belsky, D. W., Moffitt, T. E., Baker, T. B., Biddle, A. K., Evans, J. P., Harrington, H., ... & Caspi, A. (2013). Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study. *JAMA psychiatry*, 70(5), 534-542.
- Bernburg, J. G., & Krohn, M. D. (2003). Labeling, life chances, and adult crime: The direct and indirect effects of official intervention in adolescence on crime in early adulthood. *Criminology*, 41(4), 1287-1318.
- Bernburg, J. G., & Krohn, M. D. (2003). Labeling, life chances, and adult crime: The direct and indirect effects of official intervention in adolescence on crime in early adulthood. *Criminology*, 41(4), 1287-1318.
- Bernburg, J. G., & Krohn, M. D. (2003). Labeling, life chances, and adult crime: The direct and indirect effects of official intervention in adolescence on crime in early adulthood. *Criminology*, 41(4), 1287-1318.
- Bernburg, J. G., Krohn, M. D., & Rivera, C. J. (2006). Official labeling, criminal embeddedness, and subsequent delinquency: A longitudinal test of labeling theory. *Journal of Research in Crime and Delinquency*, 43(1), 67-88.
- Boardman, J. D., Blalock, C. L., & Pampel, F. C. (2010). Trends in the genetic influences on smoking. *Journal of health and social behavior*, 51(1), 108-123.
- Boardman, J. D., Blalock, C. L., & Pampel, F. C. (2010). Trends in the genetic influences on smoking. *Journal of health and social behavior*, 51(1), 108-123.
- Boardman, J. D., Blalock, C. L., & Pampel, F. C. (2010). Trends in the genetic influences on smoking. *Journal of health and social behavior*, 51(1), 108-123.
- Boffetta, P., & Hashibe, M. (2006). Alcohol and cancer. *The lancet oncology*, 7(2), 149-156.
- Botello-Harbaum, M. T., Haynie, D. L., Iannotti, R. J., Wang, J., Gase, L., & Simons-Morton, B. (2009). Tobacco control policy and adolescent cigarette smoking status in the United States. *Nicotine & Tobacco Research*, 11(7), 875-885.
- Bowditch, C. (1993). Getting rid of troublemakers: High school disciplinary procedures and the production of dropouts. *Social problems*, 40(4), 493-509.
- Brook, J. S., Balka, E. B., & Whiteman, M. (1999). The risks for late adolescence of early adolescent marijuana use. *American Journal of Public Health*, 89(10), 1549-1554.
- Bushway, S. D. (1998). The impact of an arrest on the job stability of young white American men. *Journal of research in Crime and Delinquency*, 35(4), 454-479.
- Centers for Disease Control (CDC). (1989). School Policies and Programs on Smoking and Health—United States, 1988. *MMWR. Morbidity and mortality weekly report*, 38(12), 202-203.
- Centers for Disease Control and Prevention (CDC). (1998). Response to increases in cigarette prices by race/ethnicity, income, and age groups—United States, 1976-1993. *MMWR. Morbidity and mortality weekly report*, 47(29), 605-609.
- Centola, D. (2011). An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060), 1269-1272.
- Centola, D. (2018). *How behavior spreads: The science of complex contagions* (Vol. 3). Princeton, NJ: Princeton University Press.
- Chiricos, T., Barrick, K., Bales, W., & Bontrager, S. (2007). The labeling of convicted felons and its consequences for recidivism. *Criminology*, 45(3), 547-581.
- Cuellar, A. E., & Markowitz, S. (2015). School suspension and the school-to-prison pipeline. *International Review of Law and Economics*, 43, 98-106.
- Cuellar, A. E., & Markowitz, S. (2015). School suspension and the school-to-prison pipeline. *International Review of Law and Economics*, 43, 98-106.
- DiMaggio, P., & Garip, F. (2011). How network externalities can exacerbate intergroup inequality. *American Journal of Sociology*, 116(6), 1887-1933.
- DiMaggio, P., & Garip, F. (2012). Network effects and social inequality. *Annual review of sociology*, 38, 93-118.
- Domingue, B. W., Belsky, D. W., Fletcher, J. M., Conley, D., Boardman, J. D., & Harris, K. M. (2018). The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health. *Proceedings of the National Academy of Sciences*, 115(4), 702-707.

- Domingue, B. W., Conley, D., Fletcher, J., & Boardman, J. D. (2016). Cohort effects in the genetic influence on smoking. *Behavior genetics*, 46(1), 31-42.
- Domingue, B. W., Fletcher, J., Conley, D., & Boardman, J. D. (2014). Genetic and educational assortative mating among US adults. *Proceedings of the National Academy of Sciences*, 111(22), 7996-8000.
- Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education*, 77(4), 267-282.
- Eitle, T. M., & Eitle, D. J. (2004). Inequality, segregation, and the overrepresentation of African Americans in school suspensions. *Sociological Perspectives*, 47(3), 269-287.
- Farmer, G. L. (1999). Disciplinary practices and perceptions of school safety. *Journal of Social Service Research*, 26(1), 1-38.
- Feld, S. L. (1981). The focused organization of social ties. *American journal of sociology*, 86(5), 1015-1035.
- Fletcher, J. M. (2012). Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts. *PloS one*, 7(12), e50576.
- Fletcher, J. M. (2012). Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts. *PloS one*, 7(12), e50576.
- Fletcher, J. M. (2012). Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts. *PloS one*, 7(12), e50576.
- Fletcher, J. M., Deb, P., & Sindelar, J. L. (2009). *Tobacco use, taxation and self control in adolescence* (No. w15130). National Bureau of Economic Research.
- Fowler, J. H., Settle, J. E., & Christakis, N. A. (2011). Correlated genotypes in friendship networks. *Proceedings of the National Academy of Sciences*, 108(5), 1993-1997.
- Frank, K. A., Muller, C., & Mueller, A. S. (2013). The embeddedness of adolescent friendship nominations: The formation of social capital in emergent network structures. *American Journal of Sociology*, 119(1), 216-253.
- Frank, K. A., Muller, C., Schiller, K. S., Riegle-Crumb, C., Mueller, A. S., Crosnoe, R., & Pearson, J. (2008). The social dynamics of mathematics coursetaking in high school. *American Journal of Sociology*, 113(6), 1645-1696.
- Frye, M. (2017). Cultural meanings and the aggregation of actions: The case of sex and schooling in Malawi. *American Sociological Review*, 82(5), 945-976.
- Furberg, H., Kim, Y., Dackor, J., Boerwinkle, E., Franceschini, N., Ardissino, D., ... & Tobacco and Genetics Consortium. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics*, 42(5), 441.
- Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review*, 812-828.
- Gershoff, E. T. (2010). More harm than good: A summary of scientific research on the intended and unintended effects of corporal punishment on children. *LAW & contemp. probs.*, 73, 31.
- Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Simon and Schuster.
- Haapasalo, J., & Pokela, E. (1999). Child-rearing and child abuse antecedents of criminality. *Aggression and violent behavior*, 4(1), 107-127.
- Hallinan, M. T., & Sørensen, A. B. (1985). Ability grouping and student friendships. *American Educational Research Journal*, 22(4), 485-499.
- Hecht, S. S. (1999). Tobacco smoke carcinogens and lung cancer. *JNCI: Journal of the National Cancer Institute*, 91(14), 1194-1210.
- Heitzeg, N. A. (2009). Education or Incarceration: Zero Tolerance Policies and the School to Prison Pipeline. In *Forum on public policy online* (Vol. 2009, No. 2).
- Hoffman, L. L., Hutchinson, C. J., & Reiss, E. (2009). On Improving School Climate: Reducing Reliance on Rewards and Punishment. *International Journal of Whole Schooling*, 5(1), 13-24.
- Johnson, R. A., Sotoudeh, R., & Conley, D. (2020). Polygenic scores for plasticity: A new tool for studying gene-environment interplay. *bioRxiv*.
- Kakade, M., Duarte, C. S., Liu, X., Fuller, C. J., Drucker, E., Hoven, C. W., ... & Wu, P. (2012). Adolescent substance use and other illegal behaviors and racial disparities in criminal justice system involvement: Findings from a US national survey. *American Journal of Public Health*, 102(7), 1307-1310.
- Kandel, D. B., Kiros, G. E., Schaffran, C., & Hu, M. C. (2004). Racial/ethnic differences in cigarette smoking initiation and progression to daily smoking: a multilevel analysis. *American Journal of Public Health*, 94(1), 128-135.
- Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American journal of sociology*, 115(2), 405-450.

- Kreager, D. A. (2004). Strangers in the halls: Isolation and delinquency in school networks. *Social Forces*, 83(1), 351-390.
- Kupchik, A. (2016). *The real school safety problem: The long-term consequences of harsh school punishment*. Univ of California Press.
- Kupchik, A. (2016). *The real school safety problem: The long-term consequences of harsh school punishment*. Univ of California Press.
- Lanctôt, N., Cernkovich, S. A., & Giordano, P. C. (2007). Delinquent behavior, official delinquency, and gender: Consequences for adulthood functioning and well-being. *Criminology*, 45(1), 131-157.
- Lemert, E. M. (1967). *Human deviance, social problems, and social control*. Englewood Cliffs, NJ, Prentice-Hall.
- Lemert, E. M. (1967). *Human deviance, social problems, and social control*. Englewood Cliffs, NJ, Prentice-Hall.
- Li MD, Cheng R, Ma JZ, Swan GE (2003) A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction* 98(1):23–31
- Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., ... & Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature genetics*, 51(2), 237-244.
- Lopes, G., Krohn, M. D., Lizotte, A. J., Schmidt, N. M., Vasquez, B. E., & Bernburg, J. G. (2012). Labeling and cumulative disadvantage: The impact of formal police intervention on life chances and crime during emerging adulthood. *Crime & Delinquency*, 58(3), 456-488.
- Lucas, S. R., & Berends, M. (2002). Sociodemographic diversity, correlated achievement, and de facto tracking. *Sociology of Education*, 328-348.
- Mallett, C. A. (2016). The school-to-prison pipeline: From school punishment to rehabilitative inclusion. *Preventing School Failure: Alternative Education for Children and Youth*, 60(4), 296-304.
- Mansfield, W., & Farris, E. (1992). *Public school principal survey on safe, disciplined, and drug-free schools*. Bernal Press (PA).
- Matsueda, R. L. (1992). Reflected appraisals, parental labeling, and delinquency: Specifying a symbolic interactionist theory. *American journal of sociology*, 97(6), 1577-1611.
- McFarland, D. A. (2001). Student resistance: How the formal and informal organization of classrooms facilitate everyday forms of student defiance. *American journal of Sociology*, 107(3), 612-678.
- McFarland, D. A., Moody, J., Diehl, D., Smith, J. A., & Thomas, R. J. (2014). Network ecology and adolescent social structure. *American sociological review*, 79(6), 1088-1121.
- McFarland, D. A., Moody, J., Diehl, D., Smith, J. A., & Thomas, R. J. (2014). Network ecology and adolescent social structure. *American sociological review*, 79(6), 1088-1121.
- Mizel, M. L., Miles, J. N., Pedersen, E. R., Tucker, J. S., Ewing, B. A., & D'Amico, E. J. (2016). To educate or to incarcerate: Factors in disproportionality in school discipline. *Children and youth services review*, 70, 102-111.
- Moody, J. (2001). Race, school integration, and friendship segregation in America. *American journal of Sociology*, 107(3), 679-716.
- National Center for Chronic Disease Prevention, Health Promotion (US), United States. Public Health Service. Office of the Surgeon General, National Center for Chronic Disease Prevention, & Health Promotion (US). Office on Smoking. (1998). *Tobacco Use Among United States Racial/ethnic Minority Groups*. Department of Health and Human Services, Center for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
- Nesson, E. (2017). Heterogeneity in smokers' responses to tobacco control policies. *Health economics*, 26(2), 206-225.
- Nesson, E. (2017). Heterogeneity in smokers' responses to tobacco control policies. *Health economics*, 26(2), 206-225.
- Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., ... & Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature*, 456(7218), 98-101.
- Oakes, J. (1987). Tracking in secondary schools: A contextual perspective. *Educational psychologist*, 22(2), 129-153.
- Paternoster, R., & Iovanni, L. (1989). The labeling perspective and delinquency: An elaboration of the theory and an assessment of the evidence. *Justice Quarterly*, 6(3), 359-394.
- Pentz, M. A., Brannon, B. R., Charlin, V. L., Barrett, E. J., MacKinnon, D. P., & Flay, B. R. (1989). The power of policy: the relationship of smoking policy to adolescent smoking. *American Journal of Public Health*, 79(7), 857-862.

- Perry, B. L., & Morris, E. W. (2014). Suspending progress: Collateral consequences of exclusionary punishment in public schools. *American Sociological Review*, 79(6), 1067-1087.
- Schwartz, R. D., & Skolnick, J. H. (1962). Two studies of legal stigma. *Social Problems*, 10(2), 133-142.
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. (2000). The Color of Discipline: Sources of Racial and Gender Disproportionality in School Punishment. Policy Research Report.
- Sørensen, A. B. (1970). Organizational differentiation of students and educational opportunity. *Sociology of Education*, 355-376.
- Strauss, R. S., & Pollack, H. A. (2003). Social marginalization of overweight children. *Archives of pediatrics & adolescent medicine*, 157(8), 746-752.
- U.S. Department of Health and Human Services. (2012). Preventing Tobacco Use Among Youth and Young Adults: A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
- Welch, K., & Payne, A. A. (2010). Racial threat and punitive school discipline. *Social Problems*, 57(1), 25-48.
- Westmaas, J. L., Wild, T. C., & Ferrence, R. (2002). Effects of gender in social control of smoking cessation. *Health psychology*, 21(4), 368.
- Wolf, K. C., & Kupchik, A. (2017). School suspensions and adverse experiences in adulthood. *Justice Quarterly*, 34(3), 407-430.
- Zeng, Z., & Xie, Y. (2008). A preference-opportunity-choice framework with applications to intergroup friendship. *American Journal of Sociology*, 114(3), 615-648.

Chapter 4

The Embeddedness of Adolescent Risk Behaviors

Abstract

Social scientists are increasingly considering behaviors as they happen across multiple levels of social reality. But in addition to being vertically embedded across levels of social life, behaviors are also horizontally embedded, among other behaviors. In this paper, I use relational methods frequently used in cultural sociology (Relational Class Analysis and Latent Class Analysis) to study risky behaviors among a range of other behaviors adolescents engage in, from academics to team sports to home life and leisure activities. The relational analysis reveals three types of how risky behaviors relate to other behaviors: one where risk-taking opposes all participatory activities in school including clubs, academics, and sports, and two other classes where risk-taking combines with sports and other forms of participation in school and home life. LCA finds three types of students who correspond vaguely to the three stereotypes of nerds, socially involved students (socialites), and risk-takers. Students' membership in these behavioral classes is shaped by their institutional, relational, and familial commitments as well as their demographic backgrounds. Using twin models, I also estimate the heritability of class memberships, which proves to be roughly 30% across the classes. Finally, I assess the differences in drug users' happiness, and downstream drug use in the three relational classes. The results show that adolescent drug users who are also institutionally and socially embedded have better outcomes.

Introduction

Sociologists, in their quest to understand human behavior, have tended to focus on explaining single outcomes at a time. However, a growing literature spanning medicine and social sciences argues that behaviors, attitudes, and outcomes co-occur and cluster such that it is more useful to consider them together rather than apart. For example, scholars in public health and biology increasingly consider aging as a single process with multiple different phenotypes that mark its advance, rather than as a series of separate outcomes to be treated such as hair loss or frailty. The network of physical traits (phenotypes) as they change with age, is used to inform researchers about underlying biological processes (Freund 2019). In cultural sociology, scholars argue that the meanings of attitudes are revealed through their relationship to other attitudes (DiMaggio 1997), employing relational strategies to map the cultural space of attitude and predicting positions within that space, rather than individual attitudes on their own.

This paper builds on these budding, but disparate, literatures by showing how methods from cultural sociology for the study of attitudes can be fruitfully applied to the study of human behavior to identify clusters of individuals who share patterns of behaviors. Considering patterns of behaviors in this way is not only phenomenologically more consistent with individual experience, where a wide range of behaviors are engaged in together and at times simultaneously, but can also help us model what behaviors mean, as reflected in their position vis-a-vis other behaviors. We can then model the immediate and longer term consequences of adhering to particular behavioral patterns and not others.

The case considered here is that of risk-taking in American middle and high schools. Schools are contexts in which lifelong habits are forged. Behaviors in middle and high schools

shape life trajectories, identities, and substance addictions that last beyond one's school years. Scholars have tended to focus on substance use as a distinct adolescent outcome to be explained. However, substance use is embedded in a broader ecology of behaviors and activities. What substance use means and whether it persists into adulthood depends not only on the substances themselves and the practices built around them, but also on the broader cultural and behavioral context in which they are practiced.

Advances have already been made in the public health literature, where methods such as factor analysis and Latent Class Analysis are used to understand drug use in the context of other risky behaviors. For example, scholars have identified patterns of risk-taking – which sets of drugs do students use and is drug use coupled with risky sex or criminal behavior? – and evaluated how those patterns shape future outcomes (Petridou et al 1997; Meader et al. 2016). These studies have furthered our understanding of the etiology and consequences of risk-taking and have shown that there is much to be gained from embedding risky behaviors of different domains.

The previous literature has been limited in two respects. First, behaviors are limited to only those that are deemed risky from a public health perspective. As such, they may miss the horizontal embeddedness of risky behaviors in the broader ecology of other behaviors. To give an example of why this might matter, we know from qualitative and quantitative research in the social sciences that adolescent drinking behavior is semiotically complex – it is, of course, risky for adolescents to binge drink and yet students who drink are not risk-takers per se. Drinking is common and normalized to an extent. Depending on context, students may freely consume alcohol without being stigmatized and may even gain status from doing it. If we compare alcohol only to other substances, we will struggle to understand drinking's tendency to cross social

boundaries, which is revealed in its tendency to be done by many different kinds of students in many different contexts and alongside safer behaviors. Where past papers have focused on a limited set of risky behaviors, this paper evaluates the traditional set of risky behaviors alongside a wide range of non-risky behaviors, from sports to household activities to school work. In doing so, it provides a more holistic understanding of students' relationships with alcohol and the local meaning of alcohol consumption. This holistic understanding contextualizes not only drinking but each of the behaviors which it considers.

Second, in high schools especially, and social contexts more generally, the set of behaviors that one engages in is shaped not only by one being drawn towards behaviors that one favors but also through identity-based or network-based processes which lead one to select out of, and even oppose, participation in other behaviors. It is a trope of Hollywood movies and adolescent literature that American high schools are composed of multiple and competing roles and identities – nerds, jocks, goths, and everyone in between – who define themselves through out-group exclusion as much as in-group favoritism. Outside of Hollywood, sociologists have found that American high schools are characterized by the multiplicity of identity groups and the strong boundaries that separate them (McFarland and Pals 2005). We might expect the oppositional structure of behaviors to be even more pronounced for stigmatized behaviors (which include many risky behaviors), since stigmatization can serve as the basis of identity formation (Kaufman and Johnson 2004).

Detecting opposition requires methodologies suited to the task. Methods like Latent Class Analysis (LCA) identify sets of individuals who engage in similar behaviors and sets of behaviors that share the individuals who engage in them. They are therefore useful for identifying distinct groups of people who share sets of behaviors, but have two shortcomings:

first, LCA and other similar methods do not capture whether and how these sets of behaviors relate to each other; and second, they tend to identify groups of individuals with similar *levels* of behaviors. In risky-behavior research, this generally means identifying low-, medium- and high-risk groups of individuals, which although useful, fail to identify more fundamental differences in how students engage in risky behaviors.

This paper complements a traditional LCA with Relational Class Analysis (RCA), a method recently developed in the sociology of culture. Originally developed for attitudinal analysis (Goldberg 2011), RCA seeks to identify respondents who share a similar framework for how attitudes relate to one another, independent of the attitudes that individual actually holds. When used on attitudinal data, RCA reveals similar ways of construing a given domain – for example, the organizing logic underlying one’s attitudes in the domain of politics (Baldassari and Goldberg 2014), religion and science (DiMaggio et al. 2018), economic transactions (DiMaggio and Goldberg 2018), or music (Goldberg 2011). A shared construal does not necessarily reveal shared values or attitudes; in fact, opposing attitudes or values can be part of the same RCA class. Rather, it means that respondents tend to *organize* their attitudes and values in similar ways.

An example regarding music tastes might help clarify things further. Imagine two people. Person A likes classical music, opera, and jazz but dislikes rap, heavy metal, and pop, while person B dislikes classical, opera, and jazz, but likes rap, heavy metal, and pop. Both person A and B will be assigned the same class by RCA – perhaps one with a *high-brow* vs. *low-brow* logic. While they have opposing tastes, they agree on which genres go together and which do not: classical music, opera, and jazz on one side and rap, heavy metal, and pop on the other. Other RCA classes might show different logics from this one: in the case of Goldberg (2011), in

addition to a *high vs. low brow class*, he finds a *contemporary vs. old class*, and an *omnivore class*, where respondents consume all genres. Thus, while LCA finds groups of individuals with the lowest covariance between their response patterns, RCA identifies classes with differing patterns of covariance, within which overall covariance is high but the absolute value of response pattern distances is minimized.

Extended to behaviors in high schools, RCA will find groups of students who share a similar pattern in terms of which behaviors are participated or not participated in together but who may differ in the actual behaviors they engage in. For example, we might find a class with students who engage in risky behaviors and scorn academics clustered with individuals who engage in academics but not risky behaviors. All these students share the same *academics-versus-risk* orientation organizing their behaviors, even though they engage in those behaviors differently.

RCA is therefore uniquely able to reveal broader relational structures of behavior. The structure of behaviors will reveal what a behavior means. Beyond its theoretical appeal, I show that using RCA is predictively useful as well: RCA class membership is associated with an individual's behavioral trajectory and subjective wellbeing.

Although RCA has primarily been applied to attitudes, the above lends us to think that it might be revealing for behaviors as well. However, there are also potential reasons RCA might apply poorly to behaviors. First, opposition may not characterize behavior. Strong identities in high schools, like jocks and nerds or risk-takers and cautious types, may not exist in the data. Alternatively, members of groups with distinct identities may be more similar in their overall patterns of daily activities and behavior than cultural tropes would have us expect. Identity may

be more expressed through subtle differences in dress and speech uncaptured by survey response than in behavior.

Alternatively, there may be less structure and variation in behaviors compared to attitudes if behaviors are more constrained by opportunities for action and availability of resources.

Attitudes may be more able to take on heterogeneous patternings compared to behaviors, which may be more idiosyncratic and dependent on the opportunity structure an individual encounters.

This may especially be the case for high school students, who are embedded in strong institutions, such as families and schools, which heavily shape their lives and access to behaviors. Students, for example, may have a preference for one behavior or another but not be able to express it, thus limiting the utility of relational methods that attempt to reveal frameworks through behavioral occurrence. Thus, I compare RCA results to those produced by LCA, which tends to find sets of behaviors that co-occur, and minimizes within-class covariance by design.

In addition to being horizontally embedded, behaviors are also vertically embedded in overlapping institutions and relations, including the school, the grade, friendship networks, and families. Sociologists and scholars in public health have shown that the multiple levels of high school student's social worlds shape their engagement with a wide range of behaviors, including smoking and substance use (Diez-Roux 1998; Moore et al 2018; Barnes et al 2006; Kelly et al 2011). Beyond their effects on individual behaviors, however, social contexts at different levels may also shape how and under what conditions behaviors co-occur and cohere into shared patterns and frameworks. In a series of correlational and regression-based analyses, I examine the relative contribution of contexts against individual-level factors which predict class membership. By adding information on students' genetic proclivities to the analyses, I am able to incorporate both biological and social explanations for behavioral adoption.

The combination of findings from RCA and LCA show that the behaviors of American high school students cluster into distinct sets of behavioral patterns. Both identify three classes. For RCA, the first class includes students for whom participating in major sports, solitary entertainment (TV and video games), and risky behaviors opposes participating in home and school activities, a second in which an opposition between doing drugs and participating in extracurriculars is bridged by sports and home and leisure activities, and a third in which solitary entertainment, sex, and risky behaviors oppose all other behaviors. LCA separates students into groups which I term 1) nerds, who engage in academics and solitary entertainment, 2) socialites, who engage in clubs, sports, and substance use, and 3) risk-takers, who engage in many different risky behaviors.

Students' membership in these behavioral classes is shaped by their institutional, relational, and familial commitments as well as their demographic backgrounds. Genetics strongly shape the kinds of behavioral patterns students are likely to exhibit. Through a comparison of identical and fraternal twins, I estimate that the heritability of behavioral classes (RCA and LCA) to range from 10-41%, depending on the class, in line with outcomes like Type 2 Diabetes which are known to have both genetic and social and environmental causes. Genetic proclivities towards risk-taking and education in particular influence membership into classes.

Finally, I show that considering the broader structure of behaviors can improve how we understand the outcomes and experiences of students who engage in risky behaviors. Drug users who attend schools where risky behaviors tend to be connected to sports or other activities are happier at their school than drug users who attend schools where drug use tends to be isolated from other behaviors. Similarly, drug users who are members of classes in which activities

besides risky behaviors are also done are less likely than their purely risk-taking peers to continue their drug use habits into adulthood.

Behaviors in relation

Behaviors are dually embedded: horizontally, in an ecology of other behaviors, and vertically, in nested social contexts from local networks to institutions. While sociologists have used multilevel theoretical and methodological frameworks to study the vertical embeddedness of behaviors, we have only just begun to consider the complexity that horizontal embeddedness entails.

Understanding horizontal embeddedness starts from the recognition that behaviors, like attitudes, do not occur independently. Rather, behaviors, due to a host of structural, cultural, and causal reasons, co-occur (or not) with regularity. The key insight of this paper is that the presence or lack of co-occurrence reveals modalities of experience, which have durable effects on individuals' social lives.

A robust literature in cultural sociology argues that understanding behaviors and mapping their meaning requires attending to their co-occurrence with other behaviors. Bourdieu in *Distinction*, for example, shows how cultural and economic class differences are reflected in the positions that different occupational groups take in the space of cultural products they consume and activities they pursue (Bourdieu 1987). Similarly, scholars who study institutional logics have argued that meaning adheres in the link between material practices and symbolic meanings (Friedland and Alford; Friedland 2013; Mohr and Duquenne). An important study of intimate practices among university students has shown, for example, that physical, verbal, emotional,

categorical, and moral practices engaged in with intimate others cohere into distinct dimensions that reflect the underlying institutional logics of love (Friedland et al 2014).

The potential set of mechanisms for behavioral co-occurrence are many, but it is worth mentioning a few that have been studied in sociology. First, behaviors may become correlated through network dynamics. Political sociologists, for example, have shown that feedback loops between homophily and social influence can lead to behavioral divergences between groups of people who inhabit different parts of a network structure (Dellaposta et al. 2015). Under such a framework, behavioral correlations reflect the group structures of networks. Theories from cultural sociology have also focused on implicit, cognitive frameworks such as scripts and logics, which influence how behaviors are seen and specifically, the set of other behaviors they go or do not go with (Swidler 1986; DiMaggio 1997; Frye 2017; Frye and Trinitapoli 2015; Strand and Lizardo 2015; Harding 2003). Relatedly, self-identities influence the co-occurrence of behaviors. For example, studies in symbolic interactionism have argued that individuals seek to project identities and select sets of behaviors which reinforce their identity for others (Mead 1934; Serpe and Stryker 2011). Fourth, structural theories of behavior argue that social positions, such as race, class or gender, make behaviors possible or not. Under such theories, constraints in the resources or the information environment which members of different social positions have access to will shape the actionability of behaviors and in turn, the correlation between behaviors and social positions as well as the correlation between different behaviors within social positions (Blum et al 2000; Darling-Hammond 1995; Meier et al. 2018). Finally, the burgeoning literature on sociogenomics points to another possible theory – that behavioral correlations may, at least in part, emerge thanks to the same or similar sets of genes inducing those behaviors, or what has

been called pleiotropy (Bulik-Sullivan et al. 2015; Boardman et al 2015; Fletcher and Conley 2013).

The case for embedding risky behaviors horizontally

With respect to risky behaviors, most work has examined only their co-occurrence with other risky behaviors. For example, scholars have used Latent Class Analysis (LCA) to identify sets of risky behaviors which are commonly done together and shown that individual's social demographics predict the patterns with which they engage in risk. Relatedly, the literature on gateway drugs has examined whether using so-called soft drugs (such as nicotine) may prime an individual to later use harder drugs, whether as a result of their pharmacologies (Levine et al 2011), the impact that early drug use has on mental health and substance-seeking behavior (Fleming et al 1989; Kendal and Yamaguchi 1993), context-based mechanisms through which students who use softer drugs befriend students who use harder drugs leading them to adopt harder drugs as well (Otten et al 2017; Oetting and Beauvais 1986), or as the previous chapter of this dissertation suggests, because soft and hard drugs are treated similarly by schools.

By focusing solely on risky behaviors, however, this literature is unable to address why behaviors prove to be risky for some students and not others. Take sex or drinking, for example. Sex in the context of a stable relationship, and alongside adequate contraception and protection use, is not necessarily risky. Similarly drinking can have different long-term effects and be experienced very differently depending on the context in which it is done, at parties, clubs, in small gatherings of friends, or in solitude. The experience and consequences of drinking also depend on the kind of student who is doing it: students who perform well in school or those detached from home and school life. Any given behavior might have different connotations

depending on the group structure and broader relational ecology in which it is embedded in this way.

Relatedly, focusing only on risky behaviors also prevents us from seeing the group processes that shape who partakes in them and the patterns in which they are done. Social identity theory in psychology has highlighted the importance of intergroup and specifically out-group relations, in shaping group identity and behavior (Tajfel 1974; Tajfel et al. 1979). For example, experiments have shown that in-group favoritism increases where the out-group poses a threat compared to when it is irrelevant (Turner et al 1979). Willis' (1997) ethnography of schoolboys is another example of how in- and out-group identifications can shape deviance. The *lads* in Willis' study, a group of young boys with disruptive and deviant behaviors who would soon drop out of school, juxtaposed their identity with other groups, namely *ear'oles*, who were conforming and studious students. Under such conditions, behaviors may prove less elastic, making them less likely to occur with behaviors commonly done by a salient out-group. Exploring the broader position of risky behaviors in the ecology of behaviors allows us to examine the conditions under which risky behaviors become risky and when students become more than just drinkers or students who have sex, but rather, risk-takers.

Linking vertical and horizontal embeddedness

Previous work in sociology has shown that schools, families and friend networks all play an important role in shaping adolescent behavior, and especially risky behavior (McCubbin et al. 1985; Maxwell 2002; Beal et al. 2001). Multilevel modeling in particular has been used to capture the relative contributions of each layer, as well as interactions between them, in influencing behavioral adoption. For example, Kim and Chun (2018) show that friendship with

delinquent peers, parental tendency to monitor their children, and school-level antismoking policies are all significant predictors of smoking behavior among adolescents. Domingue et al. (2018) show that some of the genetic correlation observed in peers is due to macrostructural factors which bring peers with similar genetic tendencies into affiliation.

Less work, however, explores the link between vertical and horizontal embeddedness, or how the meso-level patterns of multi-behavioral participation are shaped by social context. The majority of studies that do consider their link focus on friendship networks, finding that friends behave more similarly to one another than to other same-grade peers. In a meta-analysis, Sawka et al. (2013) show that individuals change their physical and sedentary behaviors to match that of their friends. Similarly, Haynie (2001) finds that friends influence each other with respect to delinquency, defined according to their engagement in a number of delinquent behaviors.

Beyond friendship networks, students are embedded in schools and families, which also shape the kinds of behaviors available to them and the patterns in which they engage in them. As shown by the previous chapter of this dissertation, by punishing risky behaviors schools may unintentionally bring behaviors into greater association. Schools are also contexts where relationships and identity groups form, potentially driving students to adopt certain behaviors and not others or providing opportunities to some students and not others. Perhaps the work that most closely aligns with the goals of this paper is that of Frank et al. (2008), who employ a pioneering approach in their study of patterns of course-taking among adolescents. Rather than focusing on particular classes and treating them as independent predictors of future academic performance, they study the shared patterns of students' course-taking across all courses. In the process, they identify emergent contexts in which student interactions occur. Girls who share such contexts are more responsive to their same-context peers with regards to advancement in

math, than boys. This is consequential because it is responsible for a non-trivial part of the difference between girls' and boys' advancement in math. Families may similarly shape behavioral patterns. For example, public health scholars have shown that parents, by encouraging their children to engage in extracurricular activities, leave their children with less time to spend time with friends in unsupervised contexts, reducing the likelihood that they engage in risky behaviors (Kao and Salerno, 2014).

Behaviors versus attitudes

With these different considerations in mind, this paper will seek to understand how horizontal embeddedness of behaviors is shaped by social context and how differences in how behaviors are embedded shapes their effects on downstream outcomes. To do so, it proposes modeling the horizontal embeddedness of behaviors using two separate but related methods—Latent Class Analysis (LCA) and Relational Class Analysis (LCA)—now common to cultural sociology for the study of attitudes.

Can relational methods developed by cultural sociologists for attitude data help us better understand behaviors? Behaviors and attitudes differ in important ways and it is worth considering those differences before simply porting methods for the study of one to the study of the other. RCA, in particular, was developed with a specific theory of how attitude relations might reveal schematic similarities (Boutyline 2017), shared understandings (Goldberg 2011) or more recently construals (DiMaggio and Goldberg 2018) of a given attitude domain. Whether RCA should be used for the study of behaviors depends crucially on whether we expect behaviors to exhibit the same sort of population heterogeneity in covariance patterns that attitudes have been shown to exhibit.

While dependent on cultural and cognitive processes, behaviors are enacted in the world.⁹ This makes behaviors dependent on opportunities for action, whether social, material, or cultural, which may not shape or constrain attitudes in the same way. In the case here, due to their social location in strong institutions like schools and families, adolescents may exhibit convergence on behaviors whether due to resource availability or because institutions like schools create spaces and objects (actants) which guide behaviors to certain equilibrium (Callon and Latour 1981). Even without a convergence on individual behaviors, if school and family structures are broadly similar in the kinds of behaviors they punish and reward, we may see convergence on a single RCA class, which brings the usage of RCA (or LCA for that matter) into question.¹⁰ Alternatively, if schools and families vary greatly in the types of behavior they make possible and reward, we could find as much heterogeneity as there are possibilities, such that relational methods do not contribute anything interesting, aligning with conventional sociodemographic regression models.

Is there any evidence we can marshal to evaluate whether these concerns might be valid? As mentioned previously, research, especially on risky behaviors, has shown that there are differences in the extent to which, and the patterns with which, individuals partake in behaviors. Studies have used LCA to identify different kinds of risky behavior profiles that people engage

⁹ While attitudes, as recorded through surveys, are a form of practice (by virtue of declaring the attitudes), for the purposes of this paper, I do not consider all declarative practice as behavior. I am interested in behaviors that imply physical action. Though some have questioned the reliability of surveys, and by implication behavioral self-reports (Jerolmack and Khan 2014), others have shown that data from forced-choice questionnaires (such as most items in Add Health) are more reliable than interviews. In any case, I hope that considering relations across a wide range of reported behaviors, will help address some of the reliability issues forwarded by Jerolmack and Khan (2014) (see DiMaggio, 2014).

¹⁰ A previous paper has critiqued RCA, arguing that it may find heterogeneity where there is none (Boutyline and Vaisey 2017). I use a Structural Equation Model-based strategy suggested by Boutyline (2017) to evaluate whether a multi-class solution fits the data better than a single-class solution, representing a hypothesis of population homogeneity. LCA is compatible with a set of common goodness of fit criteria (AIC and BIC), which I will use to assess its multi-class fit to the data.

in. Some of these risk profiles centered around risky sex behaviors, while for others drug use plays a more important role (Villalobos-Gallegos et al. 2019). Models with multiple LCA classes appear to fit the data better than solutions with only a single class – consistent with an assumption of population heterogeneity for behaviors (as indicated by goodness of fit criteria reported in every study that uses LCA). Beyond studies which use LCA, work across sociology has identified population heterogeneity for behavior within societies, as in the case of social roles which distinguish individuals by the set of behaviors they engage in and are expected to engage in (Merton 1949). Even within contexts as small as the family, sociologists and psychologists have shown that siblings take distinct niches and parents take on distinct behavioral roles (Sullo way 2010; Thomas 1972; Breinholt and Conley 2020).

The majority of the existing LCA studies looking at co-occurrence of risky behaviors, find different *levels* of risky behavior engagement – i.e. low-risk individuals, medium-risk individuals and high-risk individuals (Sharifi et al 2017; Gottert et al 2020; James et al 2013). While it is useful to know the different levels of risky behavior engagement for public health and policy reasons, we can learn something new by finding heterogeneity in the data according to a new relational principle of RCA. Instead of grouping people with similar levels of behaviors, RCA will detect patterns in the co-occurrence of behaviors irrespective of actual levels of engagement. This allows RCA to detect different organizing principles that govern behavioral patterning. This aspect of RCA, in addition to including different kinds of data, allows for informative patterns in the data to emerge. If risky behaviors are correlated with home activities and sports, for example, drug users might be more institutionally embedded and have access to more resources and social support, compared to a context where drug use is an isolated behavior. Non-drug users in the context are also affected by the network of behaviors. A person who plays

sports, for instance, might be more at risk of getting exposed to drugs than someone in a *separate spheres* context where drugs are isolated. Thus, the heterogeneous networks of behaviors uncovered by RCA may indeed prove useful for understanding behaviors.

Even if we find population heterogeneity and RCA and LCA are able to reveal distinct and informative class structures, studying behaviors in this way will require interpretation different from how sociologists are accustomed to interpreting relations in attitudinal data. Because behaviors are strongly constrained by the structure of opportunity, the lack of a behavior does not necessarily signal opposition towards the behavior or that an individual chose not to participate. There are many cases where students wish to partake in a behavior but are not able or allowed to - for example, they may wish to play basketball for their school but fail to make the team¹¹. Thus, behavioral patterns do not necessarily reveal a preference structure, but rather reflect the reality of which behaviors tend to covary together for a group of people.

Data

The data used in this paper comes from the National Longitudinal Survey of Adolescent to Adult Health (Add Health). Add Health is a nationally representative survey of adolescent health and behavior across 132 different middle and high schools. Students were first surveyed in 1994 and 1995. Students were surveyed in schools, where they were asked general questions about their health, academic success, membership in school activities and clubs, and friendship

¹¹ We can feel more certain about active engagement in behaviors than we can about selection out of behaviors. In classical economics literature, for example, behaviors are regarded as revealed preferences; they reveal to the researcher and other people the true intentions, held preferences, and beliefs of the actor (Samuelson 1938). According to this theory, behaviors are not only occasions for labeling and understanding others, but are also revelations of inner states (which might be reinforced through labeling by others). Though this theory rests on a set of assumptions that cannot be confirmed or falsified empirically, if we accept that behaviors are more aligned than attitudes with intentions, behavioral patterns can be indicative of something more than just structural opportunities and necessities, and incorporate individual desires and goals.

relationships. A subset of the total Add Health sample was later followed up with in their homes and asked a more detailed battery of questions. Seeking to understand the many factors that shape adolescent health, the in-home survey asked students about a number of behaviors, from substance use, to what they do at home and in their free time, to their sex behaviors. Since 1994/1995, Add Health has followed up with this in-home sample periodically, each time asking the respondents about how their lives have changed since high school, in the process adding new dimensions to an already rich data set.

The behaviors included in this study come from both the in-home and in-school portions of Add Health and include information about student's 1) participation in extracurricular clubs, 2) activities outside of school, 3) in-school activities, 4) risky behaviors and 5) romantic and friendship relationships. I tried to include activities that were not limited to institutional availability and which students across schools were reasonably able to partake in: I only included extra-curricular clubs where at least one student per school was involved in, thereby removing clubs specific to large or wealthy schools, or in which no student participated.

In addition to information about behaviors and health outcomes, Add Health also collected students' genetic data. To understand how biological proclivities affect behavioral patternings, I use summary measures of genetic tendency, known as polygenic scores (PGS), for 1) risk-taking behavior (Linner et al 2019)¹², 2) educational attainment (Lee et al. 2018), and 3) height (Yengo et al. 2018) to capture different dimensions of genetic tendency. Polygenic scores are based on independent large-scale studies (cited above) which find recurrent associations between genes and a given outcome. Each genetic locus gets an estimated effect, which is then

¹² The risk score polygenic score is based on self-reported tendency to take risks across two large data sets (the UK Biobank and 23andMe). In the UKB the question was worded as the following: "would you describe yourself as someone who takes risks?" and in 23andMe the question "In general, people often face risks when making financial, career, or other life decisions. Overall, do you feel comfortable or uncomfortable taking risks?"

multiplied by the number of risk alleles an individual has on that locus (this can be 0, 1, or 2). These effects are then summed into a score called polygenic scores. All three polygenic scores used here were constructed and made available to researchers by Add Health (Beauchamp et al. 2019; Baud and Harris 2020).

Methods

Following DiMaggio et al. (2018), I use a combination of Relational Class Analysis (RCA) and Latent Class Analysis (LCA) to map behavioral clusterings of students in the Add Health sample. LCA and RCA both seek to examine population heterogeneity by clustering survey respondents according to their responses; however, they do so, in different ways and guided by different theoretical frameworks. LCA operates under the theory that for survey responses – here, questions about respondent’s engagement in behaviors – result from a discrete, unobserved (i.e., latent) variable, which is then inferred from dependencies in the data. Since the levels of this latent variable are discrete, they can be thought of and treated as separate classes. Individuals who share a class will exhibit similar patterns of response to the survey items in question. LCA seeks to minimize variance within class, while maximizing it between classes.

LCA produces membership probabilities for the respondents, which describe the likelihood that a respondent is part of one latent class or another, and item-response probabilities, which capture the likelihood that members of a class will respond to a survey item with a given categorical or continuous value. Because LCA seeks to group respondents into classes with others who share similar patterns of response, it is conceptually like other well-known clustering algorithms such as k-means clustering, even if their underlying machinery for identifying clusters are distinct.

RCA, on the other hand, is designed to identify sets of respondents who agree on a framework for organizing survey items while potentially taking different positions with respect to those items. Thus, rather than seeking to identify latent variables which capture similar patterns of response, RCA seeks to identify shared construals or frameworks that capture which variables do or do not go together, while being indifferent to how an individual responded, positively or negatively, to each. RCA follows a general analytic structure, as outlined originally by Goldberg (2011), later clarified and refined by Boutyline (2017) and still later by Sotoudeh and DiMaggio (2021): first, a similarity metric is used to measure the extent to which the responses of individuals to pairs of items are similar; second, the absolute value of the similarity score is taken, such that respondents who responded in the same or opposite ways to the set of survey items receive high similarity scores, while those with orthogonal response patterns receive low scores. Finally, the absolute-valued similarity scores are mapped as a weighted network and a network clustering algorithm is run to identify respondents who share similar construals.

RCA assigns each respondent to a class but does not provide a straightforward mapping between classes and variables. Following Goldberg (2011), class-specific patterns of response are evaluated by correlating variables within each class. For example, in one class, respondents who do drugs may also do sports while those who do not use drugs, do not engage in sports. This would produce a positive correlation between drugs and sports. This correlation structure can in turn be compared and contrasted with the correlation structure associated with other classes to understand how class membership structures the dependence between variables.

RCA and LCA complement each other in important ways (DiMaggio et al. 2018). For example, adding LCA to a standard RCA analysis can help with the interpretation of the

composition of RCA classes. RCA classes are often made up of competing groups; for example, Republicans vs. Democrats in political settings or jocks vs. nerds, in the case examined here. This makes straightforward explorations of the demographics that compose classes difficult to interpret. For example, a non-significant coefficient for race predicting class membership, could be because race has no relationship with class membership or because classes are composed equally of white and black respondents, who turn out to hold opposing patterns of behaviors.

Conversely, adding RCA to a standard LCA routine allows one to examine whether there are broader structures that might relate latent classes to one another in the data. For example, in an analysis of political attitudes, LCA will uncover classes that accord roughly with political categories such as Republican and Democrat, but will not be able to determine if these classes are opposed. RCA, on the other hand, might uncover a class of partisans, and other classes of those whose political beliefs are organized according to alternative principles (e.g. Baldassari and Goldberg 2014). However, opposition is a strong assumption about how attitudes and, in this case, behaviors operate. It is therefore important to evaluate whether splitting the data into classes detected by RCA improves our understanding of the data. The methodological toolkit that has emerged around RCA allows us to do so.

Selecting the right distance measure for RCA

In the case of RCA, there are multiple options available for the distance measure to use. As previous work has argued (Sotoudeh and DiMaggio 2021; Grimmer and King 2009), there is no one best measure and different measures are suited to different kinds of data. Thus, Sotoudeh and DiMaggio (2021) propose a paradigm where meta-features of the data at hand are entered into a predictive model trained on a simulation to assess the relative performance of the different

measures. When applied to the present data, three distance measures were predicted to perform best: correlation, a new version of relationality developed by the authors of that paper and which uses a recursive algorithm to improve its performance, and an ensemble classifier. Of the three, I select correlation for computational reasons. Specifically, the computational complexities of currently available versions of recursive relationality and the ensemble classifier scale superlinearly with the number of respondents whereas correlation scales linearly. Given the large size of Add Health's in-home sample ($N \sim 20,000$), only correlation converges in reasonable time. This leads me to use correlation as a distance measure for RCA in this paper.

Relational Class Analysis Results

RCA sorts the data into three classes describing distinct frameworks for how behaviors in high schools fit together. Following Boutyline and Vaisey (2017), I use structural equation modeling (SEM) to compare the three-group clustering solution identified by RCA to a baseline model in which all respondents derive from a single group. Results show that the three-class solution offers a substantial improvement over the single group solution (3-group AIC: 570531; 1-group AIC: 628771.9).

I present the results using a combination of heatmaps and network plots, which visualize how the behaviors correlate within classes. Strong positive correlations signal that individuals who do one behavior in a class also tend to do the other; strong negative correlations signal the opposite. Individuals within classes do not necessarily do all the same behaviors; instead, what they share is a broader patterning of which behaviors go together. The heatmaps visualize the correlations between behaviors directly with blue representing positive correlations and red representing negative correlations and white representing no correlation. The networks, on the

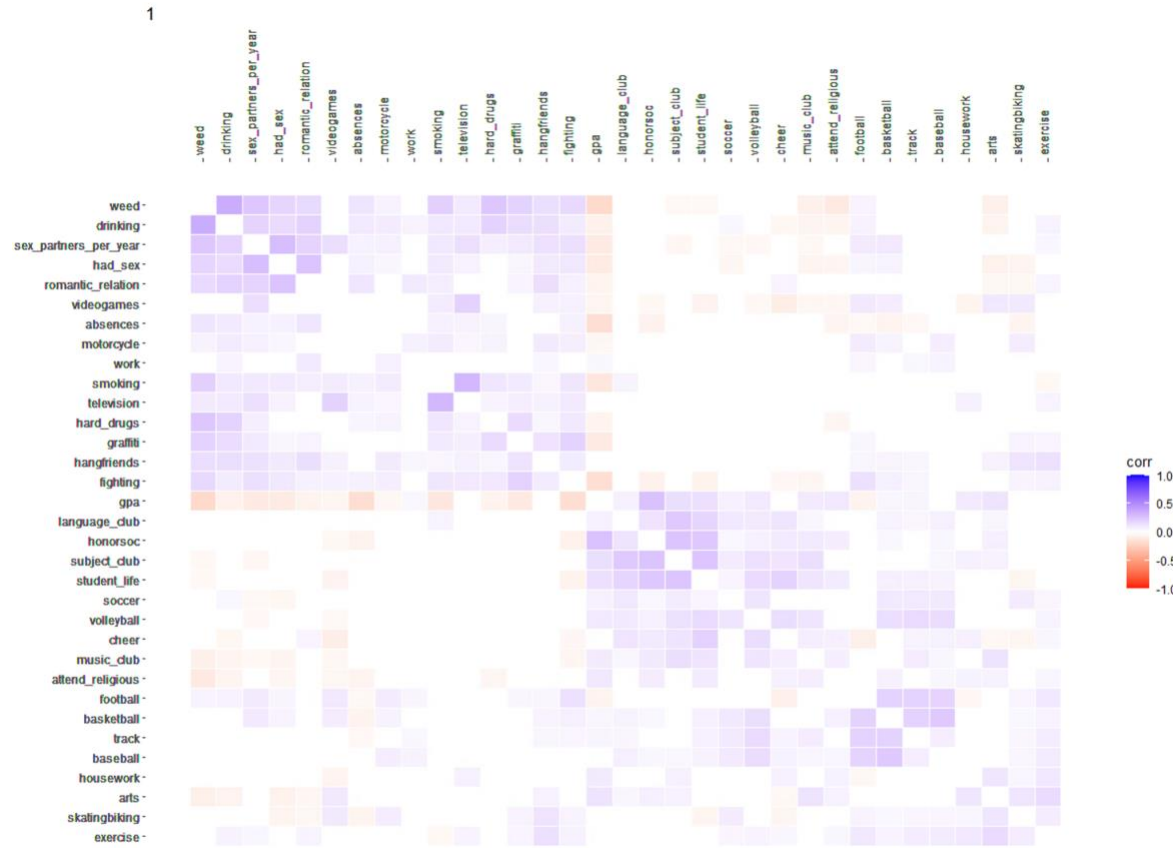
other hand, summarize the heatmaps. Behaviors were clustered qualitatively into categories based on how they tend to move together between the various classes (more information on how these categories were made can be found in Supplementary Information for Chapter 4). Nodes in the networks accord with these categories, and edges in the networks represent the average correlation of behaviors in one category with behaviors in a second category. For example, a blue-colored edge between risky behaviors and entertainment would signal that correlations between behaviors in the risky behaviors and entertainment categories are positive on average. The resulting visualizations can be found in Figure 1.

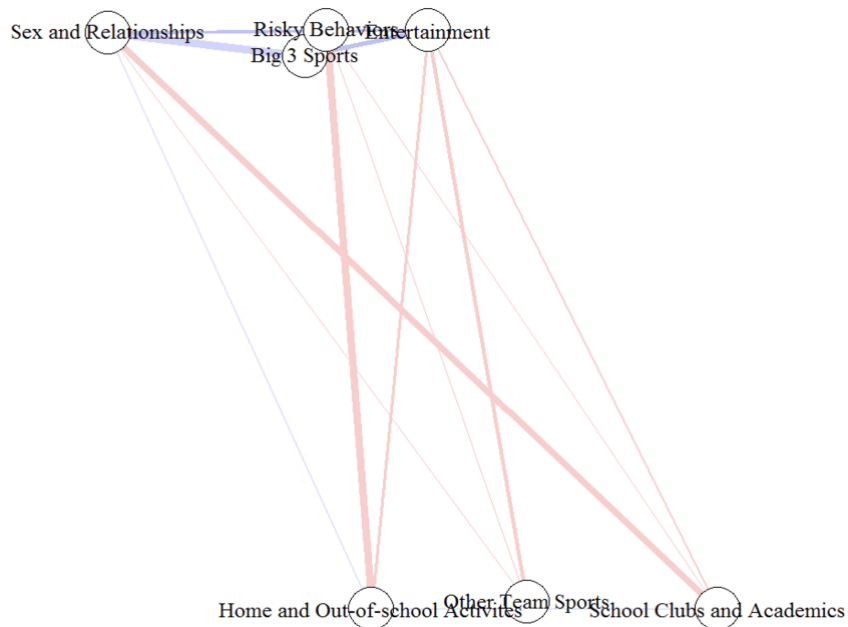
In RCA class 1, risky behaviors, sex, and entertainment combine with the big three sports (American football, baseball, and basketball) to oppose clubs, other sports and home leisure activities. RCA class 2 shows the least polarized heat map of the group. It maintains a general opposition between risky and sex behaviors and school clubs, but all sports as well as home and leisure activities bridge the two poles. Entertainment appears fairly independent from this core opposition. Finally, RCA class 3 opposes risky and sex behaviors and entertainment against all other behaviors. With respect to size, class 3 is the largest with 4788 students, while classes 1 and 2 are comparable in size with 3246 and 3260 students respectively.

Across the three classes, risky and sex behaviors oppose participation in clubs and academics. What changes are the positions of the other categories—entertainment, home and leisure, and the big three sports and other team sports. In class 1, those who engage in risky and sex behaviors do not partake in any home activities or clubs; as we move towards class 3, involvement in home and school life bridges behavioral patterns with the students involved in school life. As I will show, these modalities of behavior marked by differing degrees of institutional participation influence students' outcomes.

Figure 1: RCA heatmaps and networks

Class 1: Major sports, entertainment, and risky behaviors oppose home and school activities

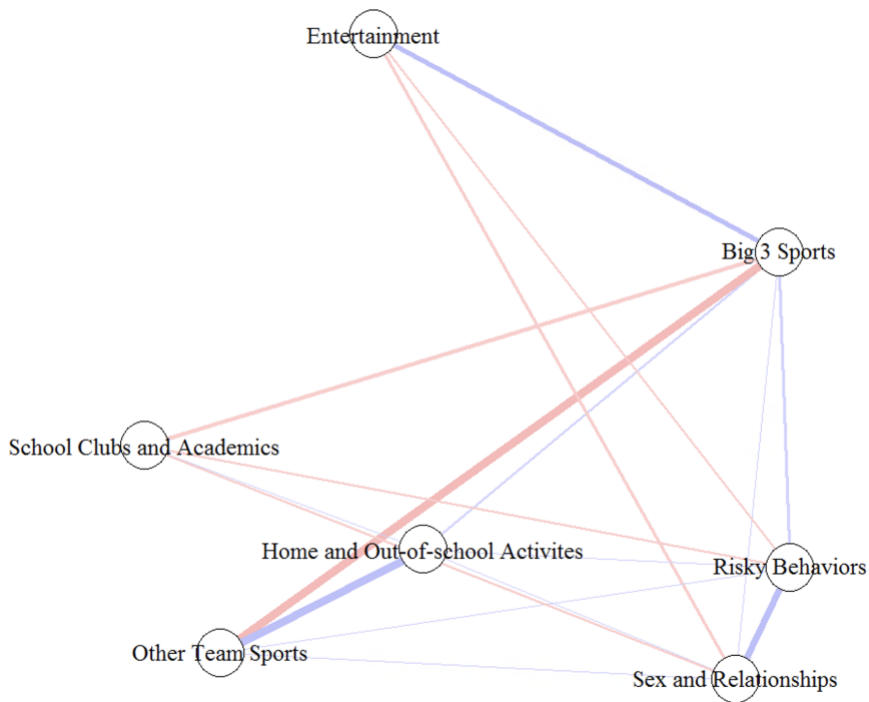




Class 2: Sports and leisure bridge drugs and extracurriculars

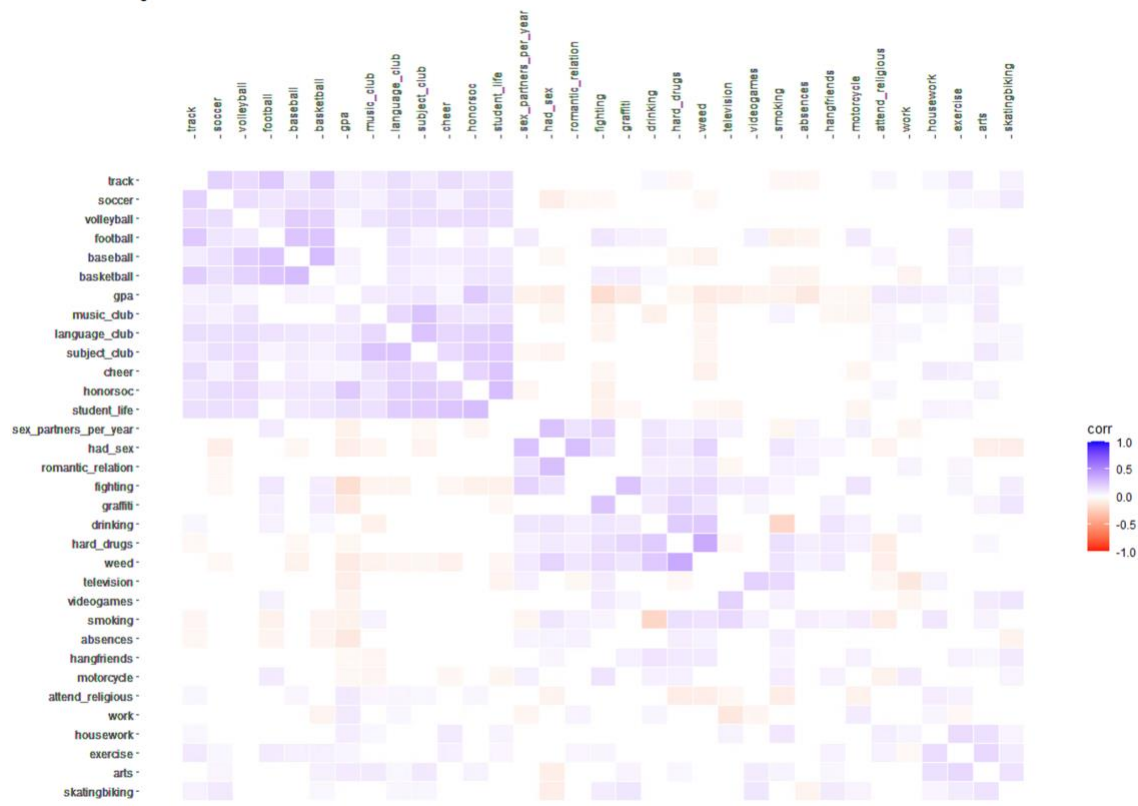
2

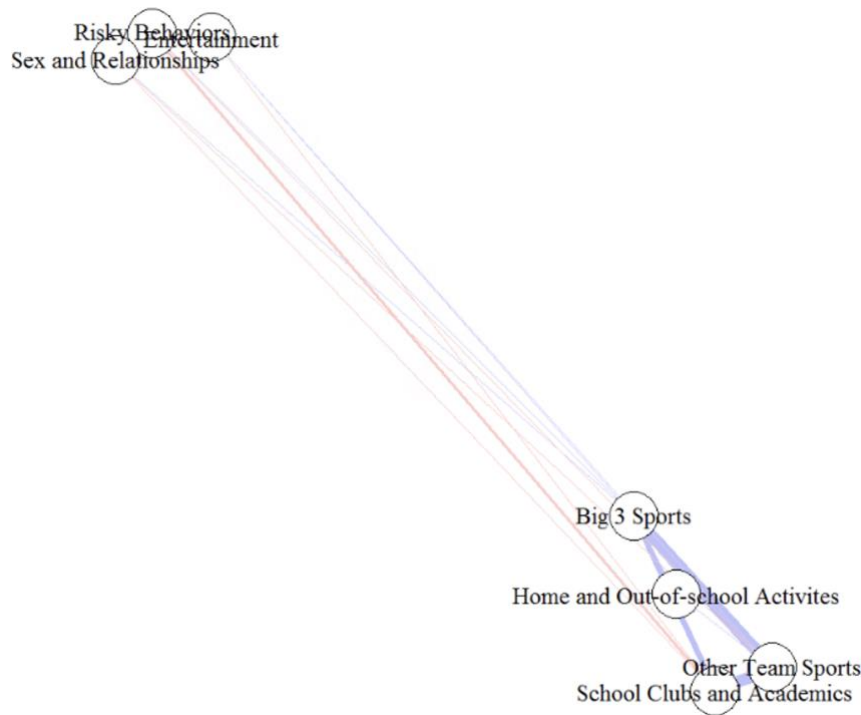




Class 3: Risky behaviors and entertainment oppose organizational engagements

3





Latent Class Analysis Results

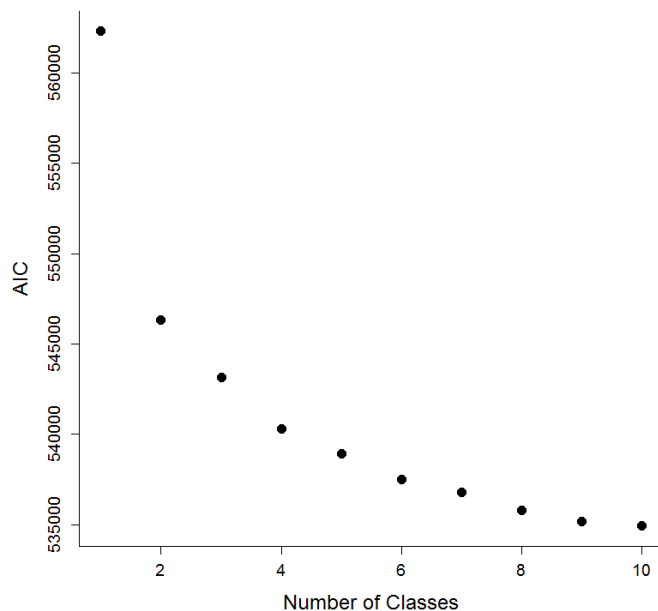
Because members of the same RCA class will likely be heterogeneous, and perhaps even opposing in their behaviors, comparing their demographic characteristics will be, at best, uninformative and, at worst, misleading. For this reason, DiMaggio et al. (2018) suggest using LCA in conjunction with RCA. LCA identifies groups of individuals who engage in similar sets of behaviors to similar degrees. As such, LCA classes can be interpreted and compared with respect to their demographics.

When running LCA, the researcher must indicate how many classes the algorithm should look for in the data. Increasing the number of class solutions will improve LCA's fit but reduce its parsimoniousness. Researchers commonly plot the number of classes against model fit, measured using information criteria, and look for "elbows" in the graph, where the number of classes is low and model fit is comparable to solutions with more classes. However, there is no

guarantee that such a solution is best or most true to the substantive case. Thus, following recommendations from the literature, I complement this quantitative approach with a qualitative evaluation of each class solution before settling on a number of classes (Masyn 2013; Bonikowski and DiMaggio 2021; Bonikowski and DiMaggio 2016).

The plot that I used to compare class solutions can be found in Figure 2. The elbow in the graph falls around the two, three, and four class solutions. At that point, the number of class solutions is low and there are diminishing returns to additional classes in terms of fit. I compared the two, three, and four class solutions and chose the three-class solution because the classes that it identified were the most qualitatively coherent. The results presented in this paper, however, are substantively robust to this choice.

Figure 2: AIC by number of classes



To understand what the different classes are capturing, I compare the means of each variable across the three classes. Values colored green connote that they are significantly higher

than the sample mean, and those in red are significantly lower. In the first class, students engage in only a select set of behaviors. They have higher GPAs and are least likely of the classes to skip school. Members of this class are more likely to partake in music club, do art outside of school, attend religious services, and watch television compared to the average high school student. Finally, they partake in honor society and subject clubs (like math and science club) at average levels. These students are engaged at home and in school, but not heavily in extracurriculars or sports and not at all in risky behaviors. In classic American high school parlance, they might be thought of as nerds.

On the other hand, students in the second class are socialites, engaging in a broad set of activities and behaviors. Compared to the previous class, these students are especially social. They have high values across the different behaviors that mark engagement in social relations (friendship and romantic partners) and school activities. Of the three classes, they engage most frequently with school clubs and sports and they have the highest average GPAs of any class. Unlike the previous class, they engage in a select set of risky behaviors, including sex as well as substance use, drinking and smoking. They avoid deviant behaviors like fighting, graffiti, and riding motorcycles. Out of the three latent classes, members of class two are least likely to partake in private entertainment like television and video games as a form of leisure.

Finally, members of class three are risk-takers, marked by their engagement in risky behaviors. They partake in substance use and sex at far higher rates than members of any other class. A quarter of class three members have tried hard drugs, for example. Compared to socialites, risk-takers participate in few, if any, extra curriculars, struggle in school, and report lower engagement in home-based activities. Their only institutional engagements are with their workplace. Where class three shares risky behaviors with class two, class three shares

entertainment with class one, spending almost three hours per day watching TV or playing video games on average.

Table 1: LCA Means

	Nerds	Socialites	Risk-Takers	Sample Mean
Baseball	0.15	0.33	0.14	0.18
Basketball	-	0.31	0.14	0.19
American Football	0.1	0.16	0.14	0.13
Track	0.11	0.28	0.08	0.13
Soccer	0.06	0.19	0.04	0.07
Volleyball	-	0.2	0.03	0.08
Cheer	0.08	0.27	0.04	0.10
Honor Society	-	0.3	0.01	0.11
Student Life	0.14	0.46	0.06	0.16
Language Club	0.11	0.37	0.07	0.14
Subject Club	-	0.4	0.06	0.16
Music Club	0.24	0.36	0.09	0.21
GPA	2.96	3.13	2.39	2.79
Smoking Cigarettes	0.23	1.4	2.51	1.23
Drinking Alcohol	0.32	1.7	2.03	1.16
Smoking Weed	0.02	0.34	0.63	0.29
Using Hard Drugs	0.02	-	0.26	0.12
Fighting	0.23	0.25	0.68	0.39
Graffiti	0.04	0.1	0.25	0.12
School Days Skipped	1.43	-	1.85	1.61
Riding Motorcycles	0.17	-	0.52	0.32
Have Had Sex	0.16	0.52	0.77	0.44
Sex Partners Per Year	0.05	-	1.26	0.56
Had Romantic Relationship	0.37	0.82	0.77	0.59
Hang out with Friends	1.71	2.17	2.22	1.97
Television	2.23	1.61	2.32	2.16

Video Games	-	0.13	0.36	0.30
Exercise	1.59	1.91	1.54	1.63
Housework	-	2.17	1.92	2.02
Arts	1.61	1.68	1.24	1.49
Skating or Biking	-	-	0.49	0.55
Attend Religious Services	0.81	0.84	0.67	0.77
Working	0.51	0.72	0.62	0.58
Column N	5393	1911	3990	11294
Column Percent	0.48	0.17	0.35	1.00

Note: Values in green are significantly higher than the sample mean; values in red are significantly lower

How do the LCA and RCA classes relate to each other? Table 2 presents a cross-table of the class solutions produced by the two methods. RCA class 1 is primarily composed of nerds and relatively fewer risk-takers, hence the opposition between risky behaviors and home- and school-based activities. RCA class 2 is composed primarily of nerds, but also a sizable portion of socialites and risk-takers. The representation of all three LCA classes may explain the unpolarized nature of RCA class 2: the opposition between nerds and risk-takers is bridged by socialites. Finally, RCA class 3 is composed primarily of risk-takers and secondarily socialites, where the institutional embeddedness of socialites sets them apart from risk-takers, despite sharing an affinity for drinking and smoking.

Table 2: Cross-tabulating members of RCA and LCA classes

		RCA Class		
		Major sports, entertainment, and risky behaviors oppose home and school activities	Sports and leisure bridge drugs and extracurriculars	Risky behaviors and entertainment oppose organizational engagements
LCA Class	Nerds	2046 (63%)	3134 (65%)	212 (7%)
	Socialites	279 (9%)	946 (20%)	688 (21%)
	Risk-Takers	935 (29%)	708 (15%)	2346 (72%)

$\chi^2 = 3683^{***}$

Note: Percentages in parentheses reflect the percent of members in a given RCA class that belong to a given LCA class.

Focusing on risk-takers specifically, most appear in RCA class 3, where they are isolated from other types of behaviors and activities. However, a non-trivial proportion are in RCA classes one and two, where their risk-taking behavior is not necessarily opposed to engagement with other behaviors. Drinking, smoking, and having sex are particularly fluid in terms of their co-occurrence with non-risky behaviors, appearing in both the socialite and risk-taker LCA classes. In addition to understanding what structures and drives selection into these different classes, a goal of subsequent analyses is to understand how participating in risky behaviors as part of a class where other non-risky behaviors are also done, may have very different social and psychological effects from participating in risky behaviors where risk-taking is isolated. Institutional embeddedness may alleviate some of the social consequences of risk taking.

How social contexts shape horizontal embeddedness

Before turning to the effects of class membership, I examine the institutional and contextual factors that structure RCA and LCA class memberships. High school students are embedded in groups and institutions which shape their lives and behaviors. Most obviously, they belong to schools and grades, composed of other students, and governed by rules and regulations, which shape their identities, behaviors, and potential for engagement in formal and informal school activities. Within schools and grades, students form friendships and romantic relationships, which perhaps even more directly shape their identities and behaviors than the school's macro-structure. Finally, high school students also belong to families, with their own, and at times, competing, rules and influences. At the same time as shaping their at-home social environment, families also provide students with their genetic backgrounds, which have their own effects on student behavior.

In Table 3, I examine the extent to which these different contexts shape student behavior, and specifically their selection into LCA and RCA classes. For schools, grades, and families, I report the intraclass correlation (ICC). Conceptually, the ICC captures the extent to which variation in the outcome is explained by groups. Higher ICC values signal that observations drawn from the group will be highly correlated, while lower values signal high degrees of heterogeneity within the group. Here, the outcomes are the probabilities that a student belongs to each of the LCA or RCA classes as derived from the LCA or RCA output and groups include schools, grades, and family units (which amount to a comparison between siblings and non-siblings). For friends, I calculate assortativity, a widely used measure of network homophily, which captures the correlation of ego and alter outcomes, i.e., class membership probabilities. Finally, in addition to examining the relative influence of genetic and social factors, I use

Falconer’s method for estimating heritability through a comparison of fraternal and identical twins (the results are robust to how heritability is estimated).

Table 3: Organizational contexts shape class membership

	Correlations by Organizational Context				Heritability
	School	Grade	Friend	Sibling	Falconer's Method
LCA 1 (Nerds)	0.05***	0.09***	0.28***	0.34***	0.34*
LCA 2 (Socialites)	0.04***	0.05***	0.13***	0.28***	0.37**
LCA 3 (Risk-Takers)	0.06***	0.08***	0.27***	0.34***	0.28*
RCA 1	0.05***	0.06***	0.09***	0.25***	0.28*
RCA 2	0.03***	0.04***	0.16***	0.22***	0.10
RCA 3	0.05***	0.06***	0.27***	0.23***	0.41**

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Institutional and social contexts shape both LCA and RCA class memberships. From the school to the family, each plays a role in influencing the collection of behaviors that a student will engage in. Across the two methods, the ICC (and assortativity) values are similar, and the strength of the correlation is stronger for closer relationships – siblings are more similar in their class membership probabilities than friends, who are more similar than school-mates and grade-mates.

Family context appears particularly important for a mix of social and genetic reasons. As the results for Falconer’s method show, identical twins are more similar in their class membership than fraternal twins, signaling genetic heritability of class membership (except for RCA class 2, where there is no significant heritability). For all classes but one (RCA class 2), the degree of heritability is roughly 0.3 to 0.4, in line with behaviors that have complex genetic and

social etiologies, such as longevity and Type 2 diabetes, and below behaviors which are strongly genetic, such as height (0.8) and Type 1 diabetes (0.88) (see <http://snpedia.com/index.php/Heritability> for a list of citations of heritability scores by trait). Previous twin studies on risk-taking tendency have also found similar levels of heritability, averaging around 30% (Linner et al 2019; Cesarini et al. 2009).

Only 5% of the variance in LCA classes is explained by school, which is not surprising, given that schools often include different identity groups. Similar levels of school ICC for RCA classes, however, are more surprising. The broader structure of behaviors could plausibly be a school-wide phenomenon: the availability of clubs along with the size of the school that determines access to those clubs, the school's policies for risky behavior punishment, and other macro-level factors shared by the entire school such as neighborhood may factor into how adolescent behaviors are related. But the results show that RCA class membership in a school is relatively low. On the other hand, correlations with friends and family as well as heritability estimates, tend to be higher. There is no way to tell, however, if the higher values for friends and family are because they fall on the same side of the RCA class (i.e., the same pole), or that they have opposing patterns of behaviors.

Sociodemographic predictors of LCA class membership

To better understand the precise demographic and genetic characteristics that influence class membership, I turn to regression analysis. I focus on predicting LCA class membership since the selection into the distinct poles of RCA classes may entail potentially contradictory structural and demographic factors. Specifically, I use multilevel models with random effects at the school level to understand how individual and school-level factors influence LCA class membership

probabilities. The results are presented in Table 4. The first three columns report models for respondents from all races without genetic predictors. The second three columns include the polygenic scores for risk-taking, education, and height. In these regressions, the sample is limited to white respondents to avoid confounding due to population heterogeneity; the first four genetic principal components of the non-Hispanic white sample are also included in these models for the same reason.

Students are more likely to belong to the first LCA class (i.e., the nerd group) if they are younger, female, non-white, come from households with married parents, and are not native to the US. They come from families with lower household income but have higher Peabody Vocabulary Test (PVT) scores, which is a quick test gauging students' ability in standard vocabulary. At the school-level, they are more likely to attend larger schools in urban areas in the South of the country. Socialites are also more likely to be female, but unlike nerds, they are more likely to be non-Hispanic whites and are more prevalent in higher grades. They have high PVT scores and come from families with higher socioeconomic backgrounds. Their schools are more likely to be private and to be in rural areas of the country, particularly in the Northeast. Finally, risk-takers are older, more likely to be male, more likely to be white and Hispanic, and more likely to be US natives. They tend to come from lower education backgrounds and have the lowest PVT scores of the three groups. Risk-takers are more likely to attend public schools in suburban areas in regions other than the southern United States, particularly the West.

In the three genetic models, we find that risk-taking PGS and education PGS predict probability of membership in LCA classes 1 and 3 (nerds and risk-takers). Specifically, risk-taking PGS negatively predicts being a nerd and positively predicts being a risk-taker, while education PGS predicts being a nerd and negatively predicts being a risk-taker. Though I

expected height PGS to predict membership in the socialite class, given its strong association with athletics, height PGS does not predict membership to any of the three classes.

Table 4: Demographic and genetic predictors of LCA membership

	Prob.Nerd	Prob. Socialite	Prob. Risk-Taker	Prob.Nerd	Prob. Socialite	Prob. Risk-Taker
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Individual Characteristics</i>						
Risk-taking PGS				-0.033*** (0.009)	0.001 (0.007)	0.031*** (0.008)
Education PGS (2018)				0.193** (0.064)	0.080 (0.050)	-0.272*** (0.059)
Height PGS				-0.007 (0.012)	0.005 (0.009)	0.002 (0.011)
Age	-0.067*** (0.008)	-0.009 (0.006)	0.076*** (0.008)	-0.044* (0.019)	-0.036* (0.015)	0.080*** (0.017)
Female	0.036*** (0.009)	0.095*** (0.006)	-0.132*** (0.008)	-0.027 (0.018)	0.120*** (0.014)	-0.093*** (0.017)
Black	0.040** (0.013)	-0.016† (0.009)	-0.024* (0.012)			
Hispanic	0.008 (0.016)	-0.050*** (0.011)	0.042** (0.014)			
Other Race	0.093*** (0.017)	-0.044*** (0.012)	-0.049** (0.016)			
PVT	0.002*** (0.0004)	0.001*** (0.0003)	-0.003*** (0.0003)	0.002** (0.001)	0.002* (0.001)	-0.004*** (0.001)
Married Parents	0.091***	0.002	-0.093***	0.104***	0.003	-0.107***

	(0.011)	(0.008)	(0.010)	(0.023)	(0.018)	(0.022)
Maternal Education	0.003	0.009***	-0.013***	0.003	0.013***	-0.016***
	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)	(0.004)
US Native	-0.123***	-0.012	0.136***	0.126	-0.113†	-0.013
	(0.017)	(0.012)	(0.016)	(0.083)	(0.065)	(0.077)
Family Income	-0.022***	0.022***	-0.0002	-0.029*	0.039***	-0.011
	(0.007)	(0.005)	(0.006)	(0.014)	(0.011)	(0.013)
Genetic PC 1 (whites only)				5.115*	0.247	-5.361**
				(1.984)	(1.548)	(1.840)
Genetic PC 2 (whites only)				3.573	-3.979†	0.406
				(2.868)	(2.238)	(2.661)
Genetic PC 3 (whites only)				6.014	4.766	-10.780
				(8.833)	(6.891)	(8.193)
Genetic PC 4 (whites only)				5.025	-3.689	-1.336
				(8.648)	(6.747)	(8.022)
8th Grade	-0.030	0.057**	-0.027	-0.085†	0.102**	-0.017
	(0.027)	(0.019)	(0.025)	(0.050)	(0.039)	(0.046)
9th Grade	-0.084**	0.073***	0.011	-0.179***	0.188***	-0.009
	(0.026)	(0.019)	(0.025)	(0.054)	(0.042)	(0.050)
10th Grade	-0.103**	0.108***	-0.005	-0.224**	0.255***	-0.032
	(0.032)	(0.022)	(0.030)	(0.068)	(0.053)	(0.063)
11th Grade	-0.100**	0.120***	-0.021	-0.294***	0.307***	-0.013
	(0.038)	(0.027)	(0.035)	(0.085)	(0.066)	(0.078)
12th Grade	-0.085†	0.158***	-0.073†	-0.262**	0.371***	-0.110
	(0.045)	(0.032)	(0.042)	(0.100)	(0.078)	(0.093)

School Characteristics

School Size	0.0001**	-0.0001***	-0.00000	0.00002	-0.0001**	0.0001†
	(0.00002)	(0.00001)	(0.00002)	(0.0001)	(0.0001)	(0.0001)
Private School	0.029†	0.045***	-0.075***	0.010	0.032	-0.042

	(0.017)	(0.012)	(0.016)	(0.032)	(0.025)	(0.030)
Urbanicity:						
Suburban	-0.035**	0.008	0.027*	-0.019	-0.010	0.029
	(0.011)	(0.008)	(0.011)	(0.025)	(0.020)	(0.024)
Urbanicity: Rural	-0.059***	0.036***	0.023†	-0.034	0.062**	-0.028
	(0.014)	(0.010)	(0.013)	(0.029)	(0.023)	(0.027)
Region: Midwest	-0.005	-0.004	0.009	0.002	0.002	-0.004
	(0.015)	(0.010)	(0.014)	(0.029)	(0.023)	(0.027)
Region: South	0.040**	-0.00003	-0.040**	0.017	-0.004	-0.013
	(0.014)	(0.010)	(0.014)	(0.030)	(0.023)	(0.027)
Region: Northeast	-0.020	0.038**	-0.017	-0.016	0.058*	-0.042
	(0.016)	(0.011)	(0.015)	(0.033)	(0.025)	(0.030)
Constant	1.563***	-0.078	-0.486***	1.038***	0.247	-0.285
	(0.129)	(0.091)	(0.120)	(0.302)	(0.236)	(0.281)
Observations	8,958	8,958	8,958	2,206	2,206	2,206
R2	0.091	0.073	0.126	0.134	0.105	0.156
Adj. R2	0.089	0.070	0.123	0.125	0.095	0.146

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Do behavioral classes matter?

Thus far, I have shown that American high school students cluster into similar sets of behaviors, and that these behavioral clusters organize into broader relational frameworks for what kinds of behaviors are done and not done together. Students' membership in behavioral classes is shaped by their institutional, relational, and familial environments as well as their demographic backgrounds. With respect to heritability, genetics play a role in shaping class memberships; risk-taking and education polygenic scores, in particular influence membership into LCA classes. LCA clustered adolescents with a tendency to engage in risky behaviors into a class which I called risk-takers. As Table 2 showed, however, some portion of risk-takers appear in all three

RCA classes. In the final set of analysis, I examine whether the relations between behaviors, as captured by RCA, can help us better understand risky behaviors in schools.

I limit this set of analyses to only include drug users – students who had used hard drugs in the past 30 days¹³. Limiting the data to only drug users has several advantages. First, I can compare across individuals who partake in the same behavior across the different classes, even though RCA classes may include different kinds of individuals. Second, it speaks to adolescent risky behaviors literature and shows the utility of a complementary relational approach (RCA), in a field where other kinds of relational methods such as LCA are being used.

RCA Class 3 is the RCA class where risk taking individuals are the most isolated: risky behaviors and solitary entertainment (such as video games and watching TV) oppose other, institutionally and socially engaging behaviors. In RCA Classes 1 and 2, on the other hand, drug users and risk-takers participate in home life, sports and even clubs. This makes RCA class 3 drug users the least institutionally and socially engaged group. I therefore compare members of RCA Class 3 to students in Classes 1 and 2.

I assess the relevance of RCA class membership on two outcomes: the first is a measure of subjective well-being, whether the student is happy at school, and the second is the student's downstream drug use at wave 3, after she or he has graduated high school. I consider RCA class membership in two ways: as an individual attribute (i.e., the class an individual was assigned to), and as an attribute of the school (i.e., the proportion of a school who in RCA class 3).

Considering RCA class as a context-level variable is interesting because it provides a measure of the ecology of behaviors that others in the context produce (here, I control for the individual's own RCA class). In the school-level regressions, I include the proportion of individuals who

¹³ Hard drugs include: cocaine, glue or solvents (inhalants), LSD, PCP, ecstasy, mushrooms, speed, ice, heroin, or pills, without a doctor's prescription

engage in different kinds of behaviors (such as participate in sports, school clubs and so on) with the idea of parsing the effect of number of people in one's context who engage in a given behavior (say, sports) and the proportion of a general behavior pattern (RCA class).

Drug use is known to be associated with many sociodemographic factors, so I include several controls to ensure that the RCA classes are not simply reflecting the clustering of people with similar background characteristics. Still, there are endogeneity issues with such an approach: if we find that RCA Class 3 members are less happy than their Class 1 and 2 counterparts, we do not know if their position in the behavioral matrix is affecting their happiness or if their unhappiness is driving them to take risks. As such, the results should be interpreted purely descriptively.

Since students have less ability to select or influence the class memberships of their fellow classmates, school-level class membership effects are less vulnerable to endogeneity due to feedback between class membership and the outcome in question. That said, these results should still be interpreted cautiously since we know from the ICCs that a significant but small proportion of the variance in RCA membership is due to schools.

The results are presented in Table 5. In the first two columns, the outcome variable is whether students feel happy at their school, rated on a 1 to 5 scale from strongly disagree to strongly agree. In the second two columns, the outcome is their drug use at wave 3 – whether they have used illicit drugs in the past 30 days – seven years later. In columns 1 and 3, the main predictors are whether an individual is a member of RCA Class 3 while in columns 2 and 4, the main predictors are the proportion of students in one's school in Class 3. Again, I use multilevel models with random-effects at the school level since this allows for both school-level and

individual-level predictors to be evaluated simultaneously. For wave 3 drug use, which is a binary outcome, logistic regression is used.

Being assigned to RCA Class 3 negatively predicts happiness at school and positively predicts downstream drug use. Drug-using students whose risky behaviors are isolated from institutional and social engagements are less happy, and more likely to continue using drugs seven years later. Thus, being institutionally embedded in their high school environment may protect against the continuation of their drug habit.

Being in a school where a higher proportion of students are part of RCA Class 3 negatively predicts how happy drug users feel at school. Although the school level proportion of RCA class 3 includes abstainers from risky behaviors, the oppositional and exclusionary structural logic of these behaviors is important. Note that the school-level results are controlling for the proportion of the school who engages in extracurriculars, team sports, drugs, drinking and smoking, so the results are not being driven by how many people engage in these different kinds of behaviors. The proportion of one's classmates in RCA Class 3 does not predict future drug use, however. This may indicate that the behavioral tendencies of one's high school peers matters while one is in high school, but fades after graduation. Perhaps the current contexts in which individuals are embedded in by Wave 3 matter more.

Table 5: Class memberships predict wave 1 happiness and wave 3 drug use

	Dependent Variable			
	Happy At School		Hard Drug Use, Wave 3	
	(1)	(2)	(3)	(4)
<i>Individual Characteristics</i>				
RCA Class 3	-0.117** (0.037)	-0.108** (0.037)	0.383*** (0.100)	0.395*** (0.101)
Age	-0.037 (0.099)	-0.046 (0.098)	-0.276 (0.251)	-0.294 (0.252)
Female	-0.142*** (0.037)	-0.143*** (0.037)	-0.303** (0.094)	-0.305** (0.095)
Black	-0.029 (0.043)	-0.021 (0.043)	0.040 (0.118)	0.052 (0.119)
Hispanic	-0.119* (0.050)	-0.115* (0.050)	-0.277* (0.132)	-0.277* (0.132)
Other Race	-0.040 (0.043)	-0.034 (0.043)	-0.311* (0.126)	-0.307* (0.127)
PVT	-0.009 (0.044)	-0.009 (0.044)	0.091 (0.110)	0.091 (0.110)
Married	-0.035 (0.039)	-0.039 (0.039)	0.271** (0.102)	0.267** (0.102)
Maternal Education	-0.162*** (0.040)	-0.161*** (0.040)	0.014 (0.101)	0.015 (0.102)
US Native	-0.023 (0.041)	-0.021 (0.041)	-0.072 (0.104)	-0.075 (0.104)
Family Income	0.044 (0.040)	0.046 (0.040)	0.005 (0.099)	0.002 (0.099)
Network Degree	0.233*** (0.038)	0.233*** (0.038)	0.110 (0.095)	0.112 (0.095)
8th Grade	-0.442† (0.263)	-0.484† (0.263)	-0.520 (0.670)	-0.560 (0.674)
9th Grade	-0.495†	-0.485†	-0.219	-0.210

	(0.277)	(0.276)	(0.671)	(0.673)
10th Grade	-0.445	-0.436	0.276	0.296
	(0.320)	(0.319)	(0.786)	(0.788)
11th Grade	-0.442	-0.426	0.199	0.222
	(0.365)	(0.363)	(0.904)	(0.906)
12th Grade	-0.377	-0.354	0.297	0.343
	(0.423)	(0.421)	(1.057)	(1.060)

School Characteristics

Prop. RCA Class 3		-0.515**		-0.568
		(0.173)		(0.420)
Mean Age	0.079	0.119†	-0.210	-0.159
	(0.070)	(0.069)	(0.190)	(0.194)
Prop. Female	-0.030	-0.045	0.056	0.041
	(0.062)	(0.058)	(0.131)	(0.131)
Prop. Black	0.065	0.048	0.134	0.097
	(0.086)	(0.082)	(0.224)	(0.221)
Prop. Hispanic	0.020	0.099	0.759*	0.840*
	(0.126)	(0.123)	(0.323)	(0.328)
Prop. Other Race	0.038	-0.008	0.009	-0.037
	(0.070)	(0.066)	(0.192)	(0.193)
Mean PVT	0.013	0.024	0.582†	0.621*
	(0.121)	(0.113)	(0.305)	(0.310)
Prop. Married	-0.029	-0.003	0.379	0.388
	(0.095)	(0.090)	(0.251)	(0.243)
Mean Maternal Education	0.019	0.028	0.001	0.007
	(0.125)	(0.117)	(0.288)	(0.288)
Prop. US Native	-0.010	0.043	0.243	0.308
	(0.104)	(0.100)	(0.259)	(0.265)
Mean Family Income	-0.033	-0.051	-0.312	-0.337
	(0.117)	(0.110)	(0.266)	(0.265)
Prop. Play Sports	0.162†	0.161*	0.251	0.253
	(0.083)	(0.079)	(0.209)	(0.207)
Prop. Play in School Clubs	-0.048	-0.017	-0.286	-0.254
	(0.074)	(0.070)	(0.186)	(0.187)

Mean School Absences	-0.155*	-0.181**	-0.074	-0.102
	(0.064)	(0.061)	(0.147)	(0.149)
Mean GPA	0.012	-0.005	-0.144	-0.163
	(0.072)	(0.068)	(0.181)	(0.182)
Mean Degree	-0.044	-0.051	0.005	-0.022
	(0.078)	(0.074)	(0.184)	(0.185)
Prop. Has Done Drugs	0.098	0.105	0.030	0.035
	(0.081)	(0.077)	(0.209)	(0.208)
Prop. Who Drink	0.038	0.199*	0.080	0.240
	(0.079)	(0.092)	(0.188)	(0.225)
Prop. Who Smoke	-0.038	0.294*	0.060	0.432
	(0.101)	(0.145)	(0.242)	(0.368)
School Size	-0.00004	-0.0001	-0.105	-0.112
	(0.0003)	(0.0002)	(0.190)	(0.190)
Private School	-0.276	-0.399*	-0.027	-0.191
	(0.211)	(0.203)	(0.497)	(0.515)
Urbanicity: Suburban	-0.281†	-0.237	0.277	0.351
	(0.160)	(0.152)	(0.386)	(0.392)
Urbanicity: Rural	-0.350†	-0.261	-0.063	0.038
	(0.194)	(0.184)	(0.453)	(0.461)
Region: Midwest	-0.061	-0.032	-0.296	-0.289
	(0.179)	(0.164)	(0.406)	(0.404)
Region: South	-0.075	-0.184	0.184	0.084
	(0.213)	(0.204)	(0.507)	(0.515)
Region: Northeast	-0.110	-0.196	0.022	-0.074
	(0.209)	(0.196)	(0.470)	(0.476)
Constant	4.104***	4.113***	-1.416	-1.444†
	(0.383)	(0.375)	(0.874)	(0.877)
Observations	1,025	1,025	770	770
Log Likelihood	-1,632.164	-1,628.746	-402.248	-401.43
Akaike IC	3,354.327	3,349.493	892.496	892.86
Bayesian IC	3,576.287	3,576.385	1,096.94	1,101.95

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Discussion

This paper provides early evidence for how attending to the horizontal embeddedness of the behaviors can help us understand the complex meanings of those behaviors and to make sense of why individuals who engage in similar sets of behaviors can still experience very different outcomes. The paper contributes to a growing literature in the sociology of culture, which shows that the position of behaviors, objects, and attitudes vis-a-vis each other influences how they are understood and enacted. This literature is helping sociologists formalize and operationalize culture with important implications for how culture as a concept is included in quantitative studies of social processes. This paper shows that extending the methods and theory that have been developed so far to the study of behaviors may prove fruitful.

Here, by combining RCA and LCA, I capture regularities in behaviors among adolescents. These methods allow me to map the variety of ways in which middle and high school students engage with the world (or not), bringing us closer to the reality of adolescent behavior. Most interestingly, behaviors, like attitudes, appear to be marked by population heterogeneity correlating in systematically different ways for different subsamples of adolescents. These heterogeneous correlation structures in turn appear to shape student outcomes. For example, risk-takers, the primary group of interest for this dissertation, are spread across the three different relational classes: one where risk-taking opposes all participatory activities in school including clubs and academics and sports, and two other classes where risk-taking combines with sports and other forms of participation in school and home life. Risk-takers who are isolated from participating in school activities and sports fare worse in both happiness and Wave 3 drug use. To understand the actual behaviors students in these RCA groups are

engaging in, I use LCA – a method that clusters individuals with who engage in similar behaviors. LCA finds a group of nerds, socialites and risk-takers.

Class membership is shaped by social context. I examined the extent to which social contexts, from the school to the family, account for the variation in class memberships that I observe. I find that individuals who have closer relationships or share more intimate relational contexts are also more similar in their class memberships. School-mates, for example, are roughly 5% correlated in their class memberships, while friends and siblings are closer to 30%. Although I am unable to test this directly, a possible implication of these findings is that class membership is a result of processes that occur largely at the local level, driven by interactional dynamics among friends and rivals rather than institution-level policies or organizational factors. For example, while we may find all three classes in a given school, at a more local level groups of friends may differentiate themselves against other groups, this may result in similar classes between different friend groups (due to opposition), and again more correlation of class membership within friend groups (due to similarity). Future research will examine the network dynamics of behavioral classes.

That said, behavioral patterns are not explained fully by social context. Individuals in the same school, siblings in the same family, and friends in the same friend group still differ considerably in their class memberships. Further, individual demographic and genetic backgrounds predict class membership above and beyond school context. Through a comparison of identical and fraternal twins using Falconer's method, I find that class membership is substantially heritable. Specifically, roughly 30% of the variance in class membership is explained by genes.

Through regression analyses predicting class membership, I show that class membership is influenced by one's age, race, and gender backgrounds as well as specific genetic proclivities to take risks and attain education. The results largely confirm that of others: in general, risk-taking behaviors are strongly and negatively associated with academic performance (Diego et al 2003; Sahranavard et al. 2018; Gibbon and Gerrard 1995). Girls are more likely to be studious and less likely to be risk takers (Wolfe et al 2008; Epstein 1998), older students (Brito et al 2015; Zweig et al 2002) and those from a lower socioeconomic background are more likely to engage in risky behaviors (Blum et al 2000).

The final set of results show that horizontal embeddedness matters for outcomes. A comparison of the outcomes of drug users in RCA class 3 to drug users in the other classes shows that being institutionally engaged—participating in school activities or academics—may have a protective effect for adolescents' health. Hard drug users in RCA classes 1 and 2 were more likely to be happy and less likely to keep using drugs in the Wave 3, roughly 7 years later. This is in line with previous research, which has shown engagement in extracurricular activities lowers the probability of risky behavior engagement through increasing social connectedness (McNeely and Falci 2004) or by simply keeping adolescents busy (Kao and Salerno 2014). These results are not causal, however, and more work must be done to isolate the effects of horizontal embeddedness from other confounding factors. Creative experimental designs akin to those pioneered by Paluk, Shepherd, and Aronow (2016) for studying conflict in high schools might be used to exogenously shift how risky behaviors co-occur with other behaviors and thereby isolate such effects.

What behaviors mean is also structured by the larger social context that the students live in, which changes over time. The data used here to discern behavioral classes was collected in

1994 and 1995 and it is reasonable to expect that the changes to adolescent culture since may have reconfigured these behavioral classes. A version of this study in 2021 would have to focus on the role of social media. Further, behaviors such as playing video games which were coupled with risky behaviors and sex in RCA class 3, could well be expected to be coupled with more mainstream activities, such as leisure or home activities, in current data. Tobacco use has also changed, as cigarettes have fallen out of fashion while e-cigarettes and marijuana have become more common and socially accepted. Future work can use the methods outlined here to assess how the behaviors of youth relate today.

Future work should also address more directly the dynamics of horizontal embeddedness - how does it emerge and change over time? Here, the correlation of behaviors is a static factor estimated at one moment in time, but students are continuously engaging in behaviors and adapting their behaviors because of the other students in their social networks. These behavioral changes in turn feedback into network processes. Dynamic research designs may help us model more directly the co-evolution of social network structure with behavioral networks of the sort studied here.

References

- Baldassarri, D., & Goldberg, A. (2014). Neither ideologues nor agnostics: Alternative voters' belief system in an age of partisan politics. *American Journal of Sociology*, 120(1), 45-95.
- Barnes, G. M., Hoffman, J. H., Welte, J. W., Farrell, M. P., & Dintcheff, B. A. (2006). Effects of parental monitoring and peer deviance on substance use and delinquency. *Journal of Marriage and Family*, 68(4), 1084-1104.
- Beal, A. C., Ausiello, J., & Perrin, J. M. (2001). Social influences on health-risk behaviors among minority middle school students. *Journal of Adolescent Health*, 28(6), 474-480.
- Beauchamp, J. P., & Karlsson Linnér, R. (2019). SSGAC Risk Tolerance: GWAS and MTAG Polygenic Scores (Ver 1.0).
- Blum, R. W., Beuhring, T., Shew, M. L., Bearinger, L. H., Sieving, R. E., & Resnick, M. D. (2000). The effects of race/ethnicity, income, and family structure on adolescent risk behaviors. *American journal of public health*, 90(12), 1879.
- Boardman, J. D., Domingue, B. W., & Daw, J. (2015). What can genes tell us about the relationship between education and health?. *Social Science & Medicine*, 127, 171-180.

- Bonikowski, B. & DiMaggio, P., (2021) Mapping Culture with Latent Class Analysis: A Response to Eger and Hjerm. *Nations and Nationalism*. Forthcoming.
- Bonikowski, B., & DiMaggio, P. (2016). Varieties of American popular nationalism. *American Sociological Review*, 81(5), 949-980.
- Bourdieu, P. (1987). *Distinction: A social critique of the judgement of taste*. Harvard university press.
- Boutyline, A. (2017). Improving the measurement of shared cultural schemas with correlational class analysis: Theory and method. *Sociological Science*, 4(15), 353-393.
- Boutyline, A., & Vaisey, S. (2017). Belief network analysis: A relational approach to understanding the structure of attitudes. *American journal of sociology*, 122(5), 1371-1447.
- Braudt, D., & Harris, K. M. (2020). Polygenic scores (pgss) in the national longitudinal study of adolescent to adult health (add health)–release 2.
- Breinolt, A., & Conley, D. (2020). *Child-Driven Parenting: Differential Early Childhood Investment by Offspring Genotype* (No. w28217). National Bureau of Economic Research.
- Brito, A. L. D. S., Hardman, C. M., & Barros, M. V. G. D. (2015). Prevalence and factors associated with the co-occurrence of health risk behaviors in adolescents. *Revista Paulista de Pediatria*, 33, 423-430.
- Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., ... & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, 47(3), 291-295.
- Callon, M., & Latour, B. (1981). Unscrewing the big Leviathan: how actors macro-structure reality and how sociologists help them to do so. *Advances in social theory and methodology: Toward an integration of micro-and macro-sociologies*, 1.
- Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Genetic variation in preferences for giving and risk taking. *The Quarterly Journal of Economics*, 124(2), 809-842.
- Darling-Hammond, L. (1995). Inequality and Access to Knowledge.
- DellaPosta, D., Shi, Y., & Macy, M. (2015). Why do liberals drink lattes?. *American Journal of Sociology*, 120(5), 1473-1511.
- Diego, M. A., Field, T. M., & Sanders, C. E. (2003). Academic performance, popularity, and depression predict adolescent substance use. *Adolescence*, 38(149), 35-43.
- Diez-Roux, A. V. (1998). Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American journal of public health*, 88(2), 216-222.
- DiMaggio, P. (1997). Culture and cognition. *Annual review of sociology*, 23(1), 263-287.
- DiMaggio, P. (2014). Comment on Jerolmack and Khan, “talk is cheap” ethnography and the attitudinal fallacy. *Sociological Methods & Research*, 43(2), 232-235.
- DiMaggio, P., & Goldberg, A. (2018). Searching for Homo economicus: Variation in Americans’ construals of and attitudes toward markets. *European Journal of Sociology/Archives Européennes de Sociologie*, 59(2), 151-189.
- DiMaggio, P., Sotoudeh, R., Goldberg, A., & Shepherd, H. (2018). Culture out of attitudes: Relationality, population heterogeneity and attitudes toward science and religion in the US. *Poetics*, 68, 31-51.
- Domingue, B. W., Belsky, D. W., Fletcher, J. M., Conley, D., Boardman, J. D., & Harris, K. M. (2018). The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health. *Proceedings of the National Academy of Sciences*, 115(4), 702-707.
- Epstein, D. (1998). Real boys don’t work: “underachievement”, masculinity. *Failing boys*.
- Fleming, R., Leventhal, H., Glynn, K., & Ershler, J. (1989). The role of cigarettes in the initiation and progression of early substance use. *Addictive behaviors*, 14(3), 261-272.
- Fletcher, J. M., & Conley, D. (2013). The challenge of causal inference in gene–environment interaction research: Leveraging research designs from the social sciences. *American journal of public health*, 103(S1), S42-S45.
- Frank, K. A., Muller, C., Schiller, K. S., Riegle-Crumb, C., Mueller, A. S., Crosnoe, R., & Pearson, J. (2008). The social dynamics of mathematics coursetaking in high school. *American Journal of Sociology*, 113(6), 1645-1696.
- Freund, A. (2019). Untangling aging using dynamic, organism-level phenotypic networks. *Cell systems*, 8(3), 172-181.
- Friedland, R. (2013). God, love, and other good reasons for practice: Thinking through institutional logics. In *Institutional logics in action, part A*. Emerald Group Publishing Limited.
- Friedland, R. and Alford, R., (1991). Bringing society back in: Symbols, practices, and institutional contradictions. *The new institutionalism in organizational analysis*, 232-263.

- Friedland, R., Mohr, J. W., Roose, H., & Gardinali, P. (2014). The institutional logics of love: Measuring intimate life. *Theory and society*, 43(3-4), 333-370.
- Frye, M. (2017). Cultural meanings and the aggregation of actions: The case of sex and schooling in Malawi. *American Sociological Review*, 82(5), 945-976.
- Frye, M., & Trinitapoli, J. (2015). Ideals as anchors for relationship experiences. *American Sociological Review*, 80(3), 496-525.
- Gibbons, F. X., & Gerrard, M. (1995). Predicting young adults' health risk behavior. *Journal of personality and social psychology*, 69(3), 505.
- Goldberg, A. (2011). Mapping shared understandings using relational class analysis: The case of the cultural omnivore reexamined. *American journal of sociology*, 116(5), 1397-1436.
- Gottert, A., Pulerwitz, J., Heck, C. J., Cawood, C., & Mathur, S. (2020). Creating HIV risk profiles for men in South Africa: a latent class approach using cross-sectional survey data. *Journal of the International AIDS Society*, 23, e25518.
- Grimmer, J., & King, G. (2009). Quantitative discovery from qualitative information: A general-purpose document clustering methodology. In *APSA 2009 Toronto Meeting Paper*.
- Harding, D. J. (2003). Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *American journal of Sociology*, 109(3), 676-719.
- Haynie, D. L. (2001). Delinquent peers revisited: Does network structure matter?. *American journal of sociology*, 106(4), 1013-1057.
- James, S., McField, E. S., & Montgomery, S. B. (2013). Risk factor profiles among intravenous drug using young adults: A latent class analysis (LCA) approach. *Addictive behaviors*, 38(3), 1804-1811.
- Jerolmack, C., & Khan, S. (2014). Talk is cheap: Ethnography and the attitudinal fallacy. *Sociological methods & research*, 43(2), 178-209.
- Kandel, D., & Yamaguchi, K. (1993). From beer to crack: developmental patterns of drug involvement. *American journal of public health*, 83(6), 851-855.
- Kao, T. S. A., & Salerno, J. (2014). Keeping adolescents busy with extracurricular activities. *The Journal of School Nursing*, 30(1), 57-67.
- Kaufman, J. M., & Johnson, C. (2004). Stigmatized individuals and the process of identity. *The Sociological Quarterly*, 45(4), 807-833.
- Kelly, A. B., O'FLAHERTY, M. A. R. T. I. N., Connor, J. P., Homel, R., Toumbourou, J. W., Patton, G. C., & Williams, J. (2011). The influence of parents, siblings and peers on pre-and early-teen smoking: A multilevel model. *Drug and alcohol review*, 30(4), 381-387.
- Kim, H. H. S., & Chun, J. (2018). Analyzing multilevel factors underlying adolescent smoking behaviors: the roles of friendship network, family relations, and school environment. *Journal of School Health*, 88(6), 434-443.
- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & 23andMe Research Team. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nature genetics*, 50(8), 1112.
- Levine, A., Huang, Y., Drisaldi, B., Griffin, E. A., Pollak, D. D., Xu, S., ... & Kandel, E. R. (2011). Molecular mechanism for a gateway drug: epigenetic changes initiated by nicotine prime gene expression by cocaine. *Science translational medicine*, 3(107), 107ra109-107ra109.
- Linnér, R. K., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., ... & Beauchamp, J. P. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature genetics*, 51(2), 245-257.
- Masyn, K. E. (2013). Latent Class Analysis and Finite Mixture Modeling. *The Oxford handbook of quantitative methods*, 2, 551.
- Maxwell, K. A. (2002). Friends: The role of peer influence across adolescent risk behaviors. *Journal of Youth and adolescence*, 31(4), 267-277.
- McCubbin, H. I., Needle, R. H., & Wilson, M. (1985). Adolescent health risk behaviors: Family stress and adolescent coping as critical factors. *Family Relations*, 51-62.
- McFarland, D., & Pals, H. (2005). Motives and contexts of identity change: A case for network effects. *Social Psychology Quarterly*, 68(4), 289-315.
- McNeely, C., & Falci, C. (2004). School connectedness and the transition into and out of health-risk behavior among adolescents: A comparison of social belonging and teacher support. *The Journal of school health*, 74(7), 284.
- Mead, G. H. (1934). *Mind, self and society* (Vol. 111). University of Chicago Press.: Chicago.

- Meader, N., King, K., Moe-Byrne, T., Wright, K., Graham, H., Petticrew, M., ... & Sowden, A. J. (2016). A systematic review on the clustering and co-occurrence of multiple risk behaviours. *BMC public health*, 16(1), 1-9.
- Meier, A., Hartmann, B. S., & Larson, R. (2018). A quarter century of participation in school-based extracurricular activities: Inequalities by race, class, gender and age?. *Journal of youth and adolescence*, 47(6), 1299-1316.
- Merton, R. K. (1949). *On sociological theories of the middle range* [1949]. na.
- Mohr, J. W., & Duquenne, V. (1997). The duality of culture and practice: Poverty relief in New York City, 1888-1917. *Theory and society*, 26(2/3), 305-356.
- Moore, G. F., Cox, R., Evans, R. E., Hallingberg, B., Hawkins, J., Littlecott, H. J., ... & Murphy, S. (2018). School, peer and family relationships and adolescent substance use, subjective wellbeing and mental health symptoms in Wales: a cross sectional study. *Child indicators research*, 11(6), 1951-1965.
- Oetting, E. R., & Beauvais, F. (1986). Peer cluster theory: Drugs and the adolescent. *Journal of Counseling & Development*, 65(1), 17-22.
- Otten, R., Mun, C. J., & Dishion, T. J. (2017). The social exigencies of the gateway progression to the use of illicit drugs from adolescence into adulthood. *Addictive Behaviors*, 73, 144-150.
- Paluck, E. L., Shepherd, H., & Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3), 566-571.
- Petridou, E., Zavitsanos, X., Dessypris, N., Frangakis, C., Mandyla, M., Doxiadis, S., & Trichopoulos, D. (1997). Adolescents in high-risk trajectory: clustering of risky behavior and the origins of socioeconomic health differentials. *Preventive medicine*, 26(2), 215-219.
- Sahranavard, S., Miri, M. R., & Salehiniya, H. (2018). The relationship between self-regulation and educational performance in students. *Journal of education and health promotion*, 7.
- Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, 5(17), 61-71
- Sawka, K. J., McCormack, G. R., Nettel-Aguirre, A., Hawe, P., & Doyle-Baker, P. K. (2013). Friendship networks and physical activity and sedentary behavior among youth: a systematized review. *International Journal of Behavioral Nutrition and Physical Activity*, 10(1), 1-9.
- Serpe, R. T., & Stryker, S. (2011). The symbolic interactionist perspective and identity theory. In *Handbook of identity theory and research* (pp. 225-248). Springer, New York, NY.
- Sharifi, H., Mirzazadeh, A., Noroozi, A., Marshall, B. D., Farhoudian, A., Higgs, P., ... & Noroozi, M. (2017). Patterns of HIV risks and related factors among people who inject drugs in Kermanshah, Iran: a latent class analysis. *Journal of psychoactive drugs*, 49(1), 69-73.
- Sotoudeh and DiMaggio (2021) "Coping with Plentitude: A Computational Approach to Selecting the Right Algorithm." *Sociological Methods and Research*. Forthcoming.
- Strand, M., & Lizardo, O. (2015). Beyond world images: Belief as embodied action in the world. *Sociological Theory*, 33(1), 44-70.
- Sulloway, F. J. (2010). Why siblings are like Darwin's finches: Birth order, sibling competition, and adaptive divergence within the family. *The evolution of personality and individual differences*, 86-119.
- Swidler, A. (1986). Culture in action: Symbols and strategies. *American sociological review*, 273-286.
- Tajfel, H. (1974). Social identity and intergroup behaviour. *Social science information*, 13(2), 65-93.
- Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader*, 56(65), 9780203505984-16.
- Thomas, D. L. (1972). Role-taking and power in social psychology. *American Sociological Review*, 605-614.
- Turner, J. C., Brown, R. J., & Tajfel, H. (1979). Social comparison and group interest in ingroup favouritism. *European journal of social psychology*, 9(2), 187-204.
- Vaisey, Stephen . 2009. "Motivation and Justification: A Dual-process Model of Culture in Action." *American Journal of Sociology* 114:1675-715.
- Villalobos-Gallegos, L., Medina-Mora, M. E., Benjet, C., Ruiz-Velasco, S., Magis-Rodriguez, C., & Marin-Navarrete, R. (2019). Multidimensional patterns of sexual risk behavior and psychiatric disorders in men with substance use disorders. *Archives of Sexual Behavior*, 48(2), 599-607.
- Willis, P. (1981). *Learning to labor: How working class kids get working class jobs*. Columbia University Press
- Wolfe, D. A., Jaffe, P. G., & Crooks, C. V. (2008). *Adolescent risk behaviors*. Yale University Press.
- Yengo, L., Sidorenko, J., Kempner, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., ... & GIANT Consortium. (2018). Meta-analysis of genome-wide association studies for height and body mass index in~ 700000 individuals of European ancestry. *Human molecular genetics*, 27(20), 3641-3649.
- Zweig, J. M., Phillips, S. D., & Lindberg, L. D. (2002). Predicting adolescent profiles of risk: looking beyond demographics. *Journal of Adolescent Health*, 31(4), 343-353.

Chapter 5

Conclusion

This dissertation brings together a range of different perspectives and methods to more holistically understand why adolescents take on and continue to engage in risk behaviors. Across three different but related papers, I show that behaviors are best understood through an integrative approach, both epistemologically and empirically. Risk behaviors are situated in the many layers of contexts, and among other behaviors. Though the chapters differ in their theoretical focuses and goals, I conclude with a few summary remarks that I felt merited mention.

First, the papers in this dissertation show that pitting genes and environments (or nature and nurture) against each other is at best unhelpful, and at worst, misleading. Research in the emerging field of sociogenomics has shown important ways that our genes interact with the environment to shape outcomes. But, as Chapter 2 suggests, genes are more intertwined with environments than simply attenuating or amplifying their effects. In important ways, genes are part of the social environment. Our social environment is made up of people's political beliefs and cultural attitudes, but also their *biologies*. Environments are not just sociocultural, but metagenomic, or consisting of the genes of others, from classmates to parents and siblings. Chapter 3 highlights the environment's multi-level nature and shows that students are affected differentially by school punishment, namely their genetic tendency to smoke, affects how they forge friendships. Chapter 4 shows that genes not only affect what behavior one engages in, but also predict whole patterns of behaviors.

Second, the papers seek to understand risk behaviors' multiple dimensions. Risk behaviors, and specifically substance use, cannot be fully understood as a relationship between an individual and a substance, as studies where demographic or genetic factors are examined as the sole determinants of risk behavior would suggest. Neither can risk behaviors be understood solely through their pharmacology, or as products of institutional policies, nor still as simple products of network diffusion processes. Substance use among adolescents is a result of *all* these processes. Rather than dyadically considering each one of these links, future research might consider how they intersect and mutually construct one another. This dissertation offers an idea for how to do so.

One difficulty that this dissertation faced and that future work seeking to integrate approaches will also face is that the different processes shaping behavioral adoption and persistence are happening at once and interacting with one another. Adolescents are simultaneously being affected by their peers' metagenomic environments, by policies of their school, by their peers' patterns of risk-taking, and by their own social background and genetic tendencies. In fact, Add Health reports that more than 8,000 research studies have been based on its data, each one uncovering a small aspect of these students' lives and influences. The key to building more predictive models of behavior will be collecting dynamic data, which shares Add Health's richness but where individuals are tracked in real time as their social lives unfold. Throughout the dissertation, behaviors in Waves 3 and 4 are used as important outcomes for testing the implications of observed relational and behavioral patterns in Wave 1. Even greater temporal resolution will further assist sociologists in their quest to make sense of social complexity.

Finally, though this dissertation did not directly consider questions of causal inference and causality, it can help us think about how to measure causes and effects. The papers all share the perspective that any effect, whether of a policy, a gene, or a behavior, cannot be understood outside of, or irrespective of, relational and cultural context. This semiotic complexity makes the identification of universal or objective causal effects fraught. Rather, causal effects should be thought of as local – dependent on the environment and cultural framework in which they are estimated.

How causal effects change in response to the social environment is one of the many promising avenues of sociogenomic research. Even genes, the metonymic epitome of an individual's inherent nature, do not always influence behaviors in a straightforward manner. They may help shape individual characteristics and tendencies, but the extent to which those characteristics and tendencies end up leading to the adoption behaviors or specific outcomes happens through cultural pathways. Counterintuitively, there is a great deal to learn about the cultural and social environment through studying genes.

Supplementary Information for Chapter 2

Appendix A – Descriptive Statistics

Table A1: Descriptive statistics

	Individual		Grade		Classroom		Friend		In-degree		Out-degree	
	level	N	level	N	level	N	level	N	friend level	N	friend level	N
Cigarettes Smoked Per												
Day (mean)	1.7275	9694	1.731	9666	1.7936	3505	1.5683	4692	1.4429	3506	1.6915	3293
	(5.0975)		(1.6667)		(3.3139)		(3.7709)		(3.7351)		(3.9124)	
Smoking PGS (mean)	0	9743	0.0006	9666	0.0419	3511	0.0392	4703	0.0833	3515	0.0678	3303
	(1)		(0.6318)		(0.7721)		(0.9251)		(0.9343)		(0.9283)	
Bad Apple												
(proportion)	0.1022	9735	0.1024	9666	0.104	3510	0.1042	4702	0.1091	3514	0.1024	3302
	(0.3029)		(0.0782)		(0.2012)		(0.2454)		(0.2654)		(0.2511)	
Shining Star												
(proportion)	0.1032	9740	0.1027	9666	0.1034	3511	0.1008	4703	0.0956	3515	0.1009	3303
	(0.3042)		(0.0781)		(0.1999)		(0.2478)		(0.2531)		(0.258)	
Male (proportion)	0.4758	9743	0.4756	9666	0.4711	3511	0.4603	4703	0.4481	3515	0.4568	3303
	(0.4994)		(0.1266)		(0.3345)		(0.4085)		(0.4353)		(0.4189)	
White (proportion)	0.599	9729	0.5989	9666	0.6198	3510	0.6309	4701	0.6498	3513	0.6525	3301
	(0.4901)		(0.3532)		(0.4205)		(0.4625)		(0.4631)		(0.4624)	
Black (proportion)	0.212	9740	0.2118	9666	0.1723	3511	0.1984	4703	0.1832	3515	0.1884	3303
	(0.4088)		(0.2658)		(0.2967)		(0.3843)		(0.3762)		(0.3824)	
Hispanic (proportion)	0.1451	9729	0.1451	9666	0.1476	3510	0.1206	4701	0.116	3513	0.1119	3301
	(0.3523)		(0.2222)		(0.2829)		(0.3051)		(0.3059)		(0.3001)	
Other (proportion)	0.0804	9737	0.0809	9666	0.0949	3511	0.0928	4703	0.0888	3515	0.0903	3303
	(0.2719)		(0.1219)		(0.2024)		(0.264)		(0.2646)		(0.2666)	
Family Income (mean)	0.462	9743	0.4621	9666	0.4849	3511	0.4639	4703	0.4653	3515	0.47	3303
	(0.4257)		(0.1743)		(0.3222)		(0.3563)		(0.3528)		(0.3576)	
Maternal Education												
(mean)	5.4146	9743	5.4149	9666	5.5447	3511	5.4452	4703	5.4541	3515	5.4698	3303
	(2.2336)		(0.8979)		(1.5595)		(1.8911)		(1.97)		(1.9184)	
Older Sibling												
(proportion)	0.4916	9728	0.4916	9666	0.5058	3509	0.508	4702	0.5086	3515	0.5151	3301

	(0.5)		(0.1328)		(0.3307)		(0.406)		(0.4308)		(0.4204)	
Household Smoker												
(proportion)	0.4624	9743	0.4621	9666	0.4397	3511	0.4507	4703	0.4469	3515	0.454	3303
	(0.4659)		(0.1504)		(0.3075)		(0.3874)		(0.406)		(0.3985)	
PC 1 (mean)	0	9743	-0.0005	9666	-0.0916	3511	-0.0484	4703	-0.0881	3515	-0.077	3303
	(1)		(0.6829)		(0.7529)		(0.9346)		(0.9233)		(0.9308)	
PC 2 (mean)	0	9743	0.0008	9666	0.0596	3511	0.0088	4703	0.0043	3515	-0.0193	3303
	(1)		(0.6256)		(0.8845)		(1.0092)		(1.0204)		(0.9895)	

Notes: Standard deviations in parentheses

Appendix B – Results as visualized in the main text

The tables below correspond to the figures reported in the main text of the article in the order that they appear in the text.

Table B1: Metagenomic effects of smoking (Figures 1 and 2)

	(1)	(2)	(3)	(4)	(5)
	Grade-mates	Classmates	Friends	In-degree	Out-degree
Mean Smoking PGS	0.0525*	-0.00902	0.0316 ⁺	0.0393*	0.0231
(residualized)	(0.0243)	(0.0168)	(0.0177)	(0.0199)	(0.0254)
<i>Controls</i>					
Smoking PGS (residualized)	0.0193 ⁺	0.00792	0.0116	-0.0000797	0.0322*
	(0.0115)	(0.0149)	(0.0144)	(0.0175)	(0.0157)
Black	-0.118***	-0.102***	-0.0492*	-0.0464 ⁺	-0.0334
	(0.0162)	(0.0157)	(0.0224)	(0.0252)	(0.0257)
Hispanic	-0.0324	-0.0279	0.0101	-0.0135	0.0207
	(0.0314)	(0.0215)	(0.0368)	(0.0340)	(0.0401)
Other Race	-0.0116	-0.00798	-0.000590	0.0153	0.00244
	(0.0160)	(0.0213)	(0.0219)	(0.0244)	(0.0237)
Household Smoker	0.134***	0.166***	0.123***	0.119***	0.111***
	(0.0164)	(0.0169)	(0.0170)	(0.0179)	(0.0200)
Older Sibling	0.0242	0.0166	0.0114	0.0123	0.0117
	(0.0159)	(0.0190)	(0.0149)	(0.0171)	(0.0182)
Maternal Education	-0.00883	-0.0333 ⁺	0.000918	-0.0232	-0.00277
	(0.0261)	(0.0180)	(0.0229)	(0.0218)	(0.0263)
Male	0.0205	-0.0000544	0.0215	0.00923	0.0195
	(0.0150)	(0.0162)	(0.0163)	(0.0169)	(0.0206)
Family Income	-0.0231*	-0.0200	0.00355	0.0193	-0.00638
	(0.00997)	(0.0175)	(0.0159)	(0.0198)	(0.0109)
Proportion Black	-0.0882	0.0362	-0.0571*	-0.0719**	-0.0542 ⁺
	(0.0598)	(0.0237)	(0.0253)	(0.0250)	(0.0308)

Proportion Hispanic	-0.0572 (0.0749)	0.0105 (0.0231)	-0.0733* (0.0314)	-0.0732* (0.0289)	-0.0663* (0.0330)
Proportion Other Race	0.0127 (0.0522)	0.00543 (0.0133)	0.0120 (0.0184)	-0.0266+ (0.0151)	0.0244 (0.0235)
Proportion Household Smoker	0.00292 (0.0338)	0.0276+ (0.0167)	0.0684*** (0.0142)	0.0879*** (0.0202)	0.0477** (0.0161)
Proportion Older Sibling	0.0251 (0.0243)	-0.0225 (0.0155)	0.0140 (0.0149)	0.00955 (0.0165)	0.0220 (0.0166)
Mean Maternal Education	0.117* (0.0464)	-0.0660*** (0.0181)	-0.0511* (0.0203)	-0.0286 (0.0175)	-0.0521* (0.0257)
Proportion Male	-0.0397 (0.0285)	-0.00346 (0.0170)	-0.0129 (0.0167)	-0.0192 (0.0221)	-0.000866 (0.0159)
Mean Family Income	0.0223 (0.0593)	0.0253 (0.0243)	-0.0354** (0.0136)	-0.0502*** (0.0147)	-0.0232+ (0.0129)
Constant	-0.134+ (0.0689)	-0.443*** (0.0571)	0.0691 (0.415)	-0.471*** (0.109)	0.336 (0.540)
<i>N</i>	3853	2820	3709	2743	2609
<i>R</i> ²	0.108	0.128	0.117	0.139	0.109
adj. <i>R</i> ²	0.086	0.096	0.081	0.091	0.057

Notes: Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B2: Metagenomic effects as instrumental variable (Figure 3)

	(1)	(2)
	2SLS – First Stage	2SLS – Second Stage
	Mean Cigarettes Per Day	Cigarettes Per Day
Mean grade-mate Smoking PGS (residualized)	0.164*	
	(0.0633)	
Mean grade-mate Cigarettes Per Day		0.322***
		(0.0598)
<i>Individual Controls:</i>		
Smoking PGS (residualized)	0.0169*	0.0141
	(0.00811)	(0.0119)
Black	-0.0182*	-0.112***
	(0.00883)	(0.0159)
Hispanic	-0.00896	-0.0294
	(0.0102)	(0.0306)
Other Race	-0.00651	-0.00958
	(0.00849)	(0.0160)
Household Smoker	-0.000978	0.135***
	(0.00964)	(0.0164)
Older Sibling	0.0107	0.0206
	(0.00773)	(0.0158)
Maternal Education	0.0183*	-0.0147
	(0.00852)	(0.0260)
Male	-0.0116	0.0243+
	(0.0104)	(0.0146)
Family Income	0.00652	-0.0251*
	(0.0141)	(0.0103)
<i>Grade-level controls:</i>		
Proportion Black	-0.542**	0.0855*
	(0.177)	(0.0418)
Proportion Hispanic	-0.293	0.0370
	(0.216)	(0.0374)
Proportion Other Race	-0.0771	0.0373+
	(0.148)	(0.0216)

Proportion Household Smoker	0.138 (0.0943)	-0.0418* (0.0191)
Proportion Older Sibling	0.134* (0.0624)	-0.0183 (0.0137)
Mean Maternal Education	0.197+ (0.113)	0.0547* (0.0263)
Mean Male	-0.0718 (0.0843)	-0.0163 (0.0168)
Mean Family Income	-0.00834 (0.166)	0.0249 (0.0206)
Constant	0.707*** (0.172)	0.227*** (0.0638)
<i>N</i>	3875	3853
<i>R</i> ²	0.841	0.074
adj. <i>R</i> ²	0.837	0.051

Notes: Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B3: The effects of bad apples and shining stars on smoking outcomes (Figure 4)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Grade-Mates	Grade-Mates	Classmates	Classmates	Friends	Friends	In-degree	In-degree	Out-degree	Out-degree
Proportion Bad Apples	0.0584** (0.0210)		-0.00697 (0.0182)		0.0181 (0.0193)		0.0296 (0.0249)		0.0135 (0.0211)	
Proportion Shining Stars		-0.0149 (0.0218)		0.00326 (0.0153)		-0.0196 (0.0130)		-0.0174 (0.0143)		-0.00876 (0.0254)
<i>Controls</i>										
Smoking PGS (residualized)	0.0181 (0.0115)	0.0162 (0.0116)	0.00847 (0.0146)	0.00816 (0.0149)	0.0107 (0.0143)	0.0110 (0.0143)	-0.00112 (0.0173)	0.000253 (0.0176)	0.0316* (0.0156)	0.0315* (0.0155)
Black	-0.117*** (0.0162)	-0.117*** (0.0161)	-0.102*** (0.0157)	-0.102*** (0.0156)	-0.0488* (0.0224)	-0.0504* (0.0221)	-0.0445+ (0.0255)	-0.0468+ (0.0251)	-0.0331 (0.0255)	-0.0346 (0.0257)
Hispanic	-0.0321 (0.0312)	-0.0308 (0.0314)	-0.0280 (0.0215)	-0.0279 (0.0216)	0.00932 (0.0365)	0.00983 (0.0365)	-0.0143 (0.0339)	-0.0140 (0.0341)	0.0204 (0.0399)	0.0203 (0.0400)

Other Race	-0.0110 (0.0160)	-0.0111 (0.0160)	-0.00797 (0.0213)	-0.00795 (0.0215)	-0.00181 (0.0219)	-0.000963 (0.0219)	0.0134 (0.0240)	0.0143 (0.0242)	0.00225 (0.0238)	0.00221 (0.0238)
Household Smoker	0.135*** (0.0164)	0.135*** (0.0164)	0.166*** (0.0168)	0.166*** (0.0169)	0.124*** (0.0171)	0.123*** (0.0171)	0.119*** (0.0179)	0.119*** (0.0179)	0.111*** (0.0198)	0.110*** (0.0198)
Older Sibling	0.0247 (0.0159)	0.0239 (0.0159)	0.0167 (0.0190)	0.0167 (0.0190)	0.0116 (0.0150)	0.0111 (0.0149)	0.0127 (0.0171)	0.0122 (0.0171)	0.0116 (0.0184)	0.0114 (0.0183)
Maternal Education	-0.00923 (0.0262)	-0.00863 (0.0262)	-0.0334+ (0.0179)	-0.0334+ (0.0180)	-0.00009 (0.0229)	0.000337 (0.0228)	-0.0240 (0.0218)	-0.0246 (0.0216)	-0.00351 (0.0263)	-0.00305 (0.0262)
Male	0.0235 (0.0149)	0.0203 (0.0151)	-0.0000938 (0.0162)	-0.000209 (0.0161)	0.0220 (0.0165)	0.0212 (0.0163)	0.00933 (0.0169)	0.00923 (0.0169)	0.0193 (0.0206)	0.0192 (0.0207)
Family Income	-0.0239* (0.0100)	-0.0232* (0.00996)	-0.0201 (0.0175)	-0.0200 (0.0175)	0.00395 (0.0159)	0.00360 (0.0161)	0.0194 (0.0198)	0.0190 (0.0200)	-0.00617 (0.0109)	-0.00654 (0.0109)
Proportion Black	-0.0595 (0.0525)	-0.0601 (0.0581)	0.0362 (0.0237)	0.0360 (0.0237)	-0.0574* (0.0254)	-0.0555* (0.0250)	-0.0737** (0.0258)	-0.0712** (0.0249)	-0.0548+ (0.0312)	-0.0526+ (0.0311)
Proportion Hispanic	-0.0572 (0.0720)	-0.0282 (0.0730)	0.0101 (0.0233)	0.0105 (0.0232)	-0.0743* (0.0317)	-0.0735* (0.0314)	-0.0745* (0.0294)	-0.0733* (0.0291)	-0.0669* (0.0331)	-0.0657* (0.0329)
Proportion Other Race	0.0234 (0.0503)	0.0207 (0.0523)	0.00538 (0.0133)	0.00512 (0.0132)	0.0128 (0.0186)	0.0135 (0.0183)	-0.0263+ (0.0151)	-0.0245+ (0.0146)	0.0253 (0.0240)	0.0256 (0.0238)
Proportion Household Smoker	0.00762 (0.0336)	0.0122 (0.0349)	0.0279+ (0.0168)	0.0277+ (0.0168)	0.0689*** (0.0143)	0.0689*** (0.0143)	0.0887*** (0.0205)	0.0894*** (0.0201)	0.0481** (0.0162)	0.0489** (0.0166)
Proportion Older Sibling	0.0290 (0.0247)	0.0216 (0.0248)	-0.0225 (0.0155)	-0.0225 (0.0156)	0.0151 (0.0147)	0.0146 (0.0148)	0.0105 (0.0164)	0.00998 (0.0165)	0.0227 (0.0166)	0.0224 (0.0167)
Mean Maternal Education	0.111* (0.0486)	0.119* (0.0497)	-0.0657*** (0.0181)	-0.0658*** (0.0181)	-0.0508* (0.0202)	-0.0512* (0.0202)	-0.0292+ (0.0175)	-0.0295+ (0.0175)	-0.0518* (0.0255)	-0.0515* (0.0254)
Mean Male	-0.0212 (0.0293)	-0.0412 (0.0291)	-0.00410 (0.0167)	-0.00412 (0.0167)	-0.0135 (0.0167)	-0.0122 (0.0167)	-0.0199 (0.0223)	-0.0191 (0.0222)	-0.00101 (0.0159)	-0.000319 (0.0157)
Mean Family Income	0.0104 (0.0601)	0.0190 (0.0622)	0.0256 (0.0244)	0.0254 (0.0240)	-0.0365** (0.0132)	-0.0361** (0.0138)	-0.0499*** (0.0144)	-0.0509*** (0.0147)	-0.0244* (0.0124)	-0.0244+ (0.0129)
Constant	-0.124+ (0.0666)	-0.0984 (0.0676)	-0.447*** (0.0595)	-0.440*** (0.0581)	0.0404 (0.421)	0.0771 (0.427)	-0.512*** (0.108)	-0.531*** (0.111)	0.315 (0.543)	0.363 (0.540)
<i>N</i>	3853	3853	2820	2820	3709	3709	2743	2743	2609	2609
<i>R</i> ²	0.109	0.108	0.128	0.128	0.116	0.116	0.139	0.138	0.109	0.109

adj. R^2	0.086	0.085	0.096	0.096	0.080	0.080	0.091	0.090	0.057	0.056
------------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Notes: Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C – Metagenomic effects for whites

We include individuals of all races but control for the individual’s own race and the racial composition of the grade. The concern with this approach is that the instrument might not work as well when applied to other races because the gene discovery was conducted on individuals of European ancestry. However, other studies have shown that—unlike the case for other phenotypes¹⁴—the smoking PGSs from GWAS based on European ancestry populations predict smoking behavior in individuals of other ancestries.¹⁵ Additionally, a GWAS of individuals of African ancestry has shown overlapping SNPs with European ancestry populations.¹⁶

We also conduct additional analyses where we test whether the findings hold for same race egos and peers (i.e., individuals affected by their same race peers). We only test the within-race effects for white respondents because the number of grades with substantial non-white peers is too low to allow for this test. Principle Components were constructed on the white subsample and the first two are included in the analyses. Table C1 shows that the effect of white peers on white egos is again positive and significant at the grade-mate level, even if the effect size is attenuated from what we find for the whole sample. Classroom and friend regressions (including in- and out- degree nominations), both likely underpowered, fail to reach significance.

Table C1: Metagenomic effects for whites only

	(1)	(2)	(3)	(5)	(6)
	Grade-mates	Classmates	Friends	In-degree	Out-degree

¹⁴ Bamshad M. (2005). Genetic influences on health: does race matter? *Jama* 294:937-946.

¹⁵ Saccone N.L., et al. (2009). The CHRNA5-CHRNA3-CHRNA4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res.* 69:6848-56.

¹⁶ Chen L.S., et al. (2012). Smoking and genetic risk variation across populations of European, Asian, and African American ancestry—a meta-analysis of chromosome 15q25. *Genet Epidemiol* 36(4):340-51.

Mean Smoking PGS	0.0697*	-0.00884	0.0208	0.00764	0.0391
(residualized)	(0.0281)	(0.0247)	(0.0207)	(0.0242)	(0.0262)
<i>Controls:</i>					
Smoking PGS (residualized)	-0.00347	-0.0252	-0.00344	-0.0216	0.0333
	(0.0188)	(0.0230)	(0.0219)	(0.0269)	(0.0221)
Household Smoker	0.166***	0.199***	0.125***	0.130***	0.0914***
	(0.0209)	(0.0240)	(0.0217)	(0.0267)	(0.0245)
Older Sibling	0.0379+	0.00700	0.0220	0.0291	0.00606
	(0.0194)	(0.0284)	(0.0199)	(0.0278)	(0.0218)
Maternal Education	-0.0455*	-0.0559+	-0.0304	-0.0578*	-0.0402
	(0.0207)	(0.0290)	(0.0224)	(0.0276)	(0.0257)
Male	-0.0109	-0.0192	-0.0137	0.0000170	-0.0239
	(0.0170)	(0.0247)	(0.0196)	(0.0265)	(0.0227)
Family Income	-0.0387**	-0.0411+	0.00725	0.0374	0.0147
	(0.0135)	(0.0232)	(0.0249)	(0.0330)	(0.0239)
Proportion Household Smoker	-0.0607	0.0242	0.0642**	0.0600*	0.0435+
	(0.0499)	(0.0257)	(0.0240)	(0.0297)	(0.0263)
Proportion Older Sibling	-0.0246	-0.00361	0.000652	0.00773	-0.00146
	(0.0401)	(0.0279)	(0.0229)	(0.0262)	(0.0261)
Mean Maternal Education	0.000236	-0.127***	-0.0748**	-0.0773**	-0.0531+
	(0.0549)	(0.0306)	(0.0257)	(0.0269)	(0.0314)
Proportion Male	-0.0681	-0.0123	-0.0185	-0.0218	-0.0206
	(0.0461)	(0.0276)	(0.0267)	(0.0386)	(0.0242)
Mean Family Income	-0.0450	0.0416	-0.0538***	-0.0518**	-0.0412
	(0.0669)	(0.0283)	(0.0160)	(0.0157)	(0.0279)
Constant	-0.105	-0.341**	-0.320**	-0.271*	-0.317***
	(0.123)	(0.108)	(0.110)	(0.118)	(0.0741)
<i>N</i>	2160	1348	1805	1380	1303
<i>R</i> ²	0.127	0.122	0.148	0.159	0.144
adj. <i>R</i> ²	0.092	0.068	0.090	0.084	0.063

Notes: Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In Table C2, we replicate the findings for Table S3 on the white sample. The direction and magnitude of the coefficients tell a similar story as the results for the whole sample and again only the proportion of bad apples in one's grade is statistically significant.

Table C2: Bad apples and shining stars for whites

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Grade- Mates	Grade- Mates	Classmates	Classmates	Friends	Friends	In- degree	In- degree	Out- degree	Out- degree
Proportion Bad Apples	0.0613* (0.0288)		-0.0364+ (0.0199)		0.00828 (0.0195)		0.00703 (0.0258)		0.0287 (0.0233)	
Proportion Shining Stars		0.00302 (0.0315)		0.0189 (0.0258)		-0.0208 (0.0183)		-0.0187 (0.0212)		-0.0277 (0.0196)
<i>Controls</i>										
Smoking PGS (residualized)	-0.00566 (0.0187)	-0.00857 (0.0190)	-0.0259 (0.0228)	-0.0250 (0.0228)	-0.00440 (0.0220)	-0.00341 (0.0219)	-0.0219 (0.0268)	-0.0208 (0.0267)	0.0324 (0.0222)	0.0325 (0.0221)
Household Smoker	0.168*** (0.0210)	0.167*** (0.0212)	0.200*** (0.0240)	0.199*** (0.0240)	0.126*** (0.0216)	0.125*** (0.0216)	0.130*** (0.0267)	0.130*** (0.0267)	0.0935*** (0.0243)	0.0900*** (0.0246)
Older Sibling	0.0377+ (0.0192)	0.0378+ (0.0192)	0.00763 (0.0284)	0.00639 (0.0283)	0.0221 (0.0199)	0.0213 (0.0200)	0.0292 (0.0278)	0.0280 (0.0279)	0.00579 (0.0218)	0.00473 (0.0216)
Maternal Education	-0.0463* (0.0207)	-0.0477* (0.0206)	-0.0542+ (0.0289)	-0.0561+ (0.0291)	-0.0310 (0.0224)	-0.0303 (0.0224)	-0.0580* (0.0276)	-0.0576* (0.0274)	-0.0418 (0.0258)	-0.0404 (0.0255)
Male	-0.00901 (0.0169)	-0.00987 (0.0170)	-0.0207 (0.0246)	-0.0186 (0.0249)	-0.0132 (0.0196)	-0.0135 (0.0196)	0.000007 (0.0265)	-0.00023 (0.0265)	-0.0242 (0.0227)	-0.0237 (0.0227)
Family Income	-0.0377** (0.0134)	-0.0377** (0.0132)	-0.0414+ (0.0230)	-0.0415+ (0.0234)	0.00748 (0.0249)	0.00672 (0.0248)	0.0379 (0.0333)	0.0371 (0.0330)	0.0147 (0.0238)	0.0137 (0.0235)
Proportion Household Smoker	-0.0440 (0.0482)	-0.0557 (0.0544)	0.0262 (0.0259)	0.0251 (0.0255)	0.0652** (0.0239)	0.0637** (0.0238)	0.0602* (0.0297)	0.0593* (0.0295)	0.0450+ (0.0264)	0.0454+ (0.0268)
Proportion Older Sibling	-0.0229	-0.0198	-0.00426	-0.00440	0.00110	0.00156	0.00783	0.00808	-0.00228	-0.000577

	(0.0394)	(0.0399)	(0.0277)	(0.0278)	(0.0229)	(0.0230)	(0.0262)	(0.0264)	(0.0261)	(0.0260)
Mean Maternal Education	-0.00393	-0.0186	-0.126***	-0.126***	-0.0744**	-0.0760**	-0.0769**	-0.0780**	-0.0528 ⁺	-0.0536 ⁺
	(0.0548)	(0.0576)	(0.0307)	(0.0307)	(0.0256)	(0.0259)	(0.0270)	(0.0269)	(0.0313)	(0.0316)
Mean Male	-0.0523	-0.0598	-0.0131	-0.0116	-0.0187	-0.0183	-0.0221	-0.0216	-0.0224	-0.0199
	(0.0492)	(0.0497)	(0.0273)	(0.0278)	(0.0265)	(0.0267)	(0.0386)	(0.0386)	(0.0240)	(0.0243)
Mean Family Income	-0.0405	-0.0373	0.0458	0.0416	-0.0540***	-0.0547***	-0.0517**	-0.0524***	-0.0413	-0.0434
	(0.0660)	(0.0673)	(0.0294)	(0.0281)	(0.0161)	(0.0159)	(0.0158)	(0.0157)	(0.0288)	(0.0281)
Constant	-0.149	-0.140	-0.358***	-0.343**	-0.325**	-0.330**	-0.272*	-0.280*	-0.323***	-0.328***
	(0.127)	(0.132)	(0.104)	(0.113)	(0.113)	(0.110)	(0.119)	(0.117)	(0.0755)	(0.0734)
<i>N</i>	2160	2160	1348	1348	1805	1805	1380	1380	1303	1303
<i>R</i> ²	0.126	0.125	0.124	0.123	0.148	0.148	0.159	0.159	0.144	0.144
adj. <i>R</i> ²	0.092	0.091	0.070	0.069	0.090	0.090	0.084	0.085	0.062	0.062

Notes; Standard errors in parentheses

⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix D – Understanding the magnitude of the grade-level, compared to friend-level, coefficients

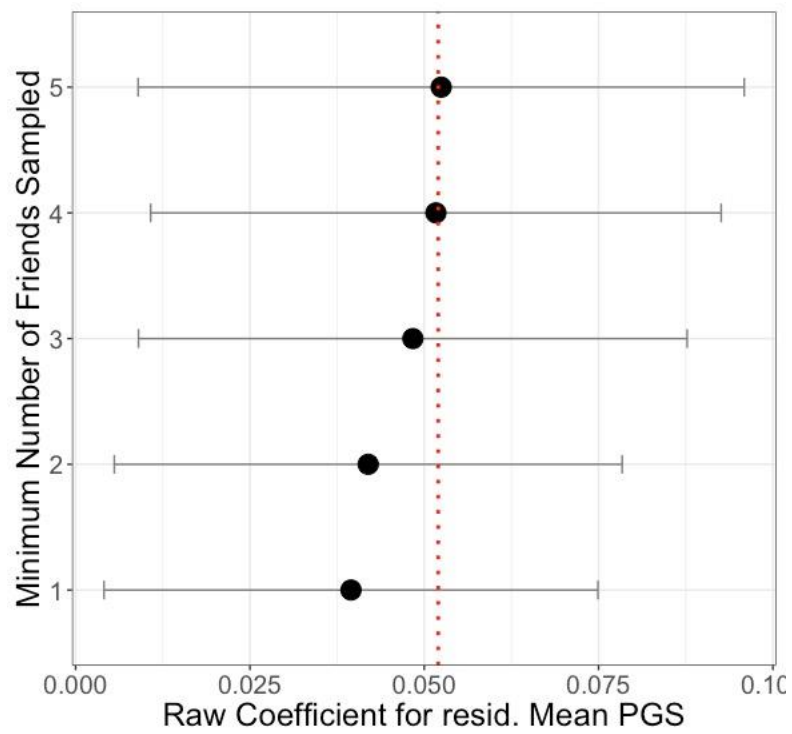
The friend-level coefficient is lower than the grade-level coefficient in the main results, despite the likely presence of homophily at the friend-level and the general assumption from sociological literature that closer relationships should be more influential than distant ones.

We hypothesized that this slight difference in coefficient size was due to the fact that friend groups were usually smaller and therefore estimates of friend-level PGS were noisier, which was compounded by the 2-to-9 sampling rate of adolescents (and thus 2-to-9 sampling of friends they nominated during the in-school interview) from the in-school sample for in-home interviews in which DNA was collected¹⁷, all resulting in a higher noise-to-signal ratio in the friend-group when compared to the grade.

We tested this hypothesis by evaluating the estimates from friend groups which included varying numbers of respondents. Since the intended outcome of this exercise was to explain both homophily and influence, we use coefficients from a naïve model that contains school and grade level fixed effects as the only controls. Figure D1 shows below that as the number of friends increases, the coefficient approaches the estimate we obtained using the grade-level models marked by the dashed red line.

¹⁷ Harris, K.M., et al. (2019). Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health). *International Journal of Epidemiology*. Forthcoming.

Figure D1: Coefficient size as a function of number of peers sampled



Supplementary Information for Chapter 3

Appendix A

In Appendix A, I examine the robustness of the results to model specification. Specifically, I evaluate two other regression-based strategies for examining the relationship between smoking punishment and smoking homophily in schools.

Whereas in the main text the relationship between punishment and homophily is modeled at the individual level, Table S1 presents results for models where *school-level smoking homophily* (measured as the correlation between ego and alter smoking behavior within schools, also known in the network's literature as assortativity) *is modeled as a function of school-level punishment (either binarized on suspension or continuous), controlling for school-level covariates*. The effect of school punishment is significant and substantively the same as those in the main text; however, because they are at the school-level they do not allow us to examine how individual-level factors, including genes, moderate the effect. Hence, I include individual-level models in the main text.

Table S1: School-level results

	Smoking Homophily (Corr.)	Smoking Homophily (Corr.)
Suspension or Harsher for First Smoking Offense	0.121*** (0.046)	
Harshness of Punishment for First Smoking Offense		0.059** (0.025)
Punishment Profile	0.039 (0.048)	0.035 (0.049)
Proportion Regular Smokers	0.179***	0.173***

	(0.067)	(0.065)
Proportion Female	-0.003 (0.330)	0.019 (0.334)
Proportion Black	0.038 (0.097)	0.038 (0.098)
Proportion Hispanic	0.117 (0.151)	0.123 (0.153)
Proportion Other Race	-0.261 (0.200)	-0.224 (0.203)
Mean Family Income	0.118 (0.229)	0.059 (0.228)
Mean Maternal Education	-0.008 (0.046)	0.014 (0.046)
Size of School Network	-0.00003 (0.0001)	-0.00002 (0.0001)
Constant	-0.185 (0.370)	-0.548 (0.388)
Observations	85	85
R2	0.174	0.160

Note: *p<0.1; **p<0.05; ***p<0.01

In Table S2, I present results from dyadic models. In the main text, the primary outcome is the proportion of one's friends who are regular smokers. Here, instead of averaging over friends, I predict directly the smoking behavior of alters. This means that there is a row in the data for each ego-alter pair, and egos and alters can each appear in the data multiple times. This is a more complex multilevel structure that demands a nested model in which dyads are nested in schools. Controls are the same as those in the main results, except that individual-level controls must be included twice: once for ego and once for alter. For the sake of brevity, the effect sizes and standard errors for controls are not shown but are available upon request (they align with those in the main text). The dyadic results for the core homophily models as well as the genetic models replicate those in the main text.

Table S2: Dyadic level results:

	All respondents (In-school sample)	Alter's Smoking Behavior	
		Genetic sample (All)	Genetic sample (Whites only)
Smoking (Ego)	0.083*** (0.003)	0.128** (0.039)	-0.007 (0.068)
Smoking Punishment (First Offense)	-0.007 (0.007)	-0.058 (0.083)	0.004 (0.128)
Smoking PGS (Ego)		0.052 (0.062)	0.176† (0.095)
Smoking (Ego) x Smoking Punishment (First Offense)	0.008* (0.003)	0.063 (0.074)	0.223 (0.145)
Smoking (Ego) x Smoking PGS (Ego)		0.375*** (0.027)	-0.112 (0.137)
Smoking Punishment (First Offense) x Smoking PGS (Ego)		0.070 (0.053)	0.003 (0.100)
Smoking (Ego) x Smoking Punishment (First Offense) x Smoking PGS (Ego)		-0.155* (0.072)	-0.120 (0.172)
Constant	-0.028† (0.017)	0.004 (0.189)	0.637 (0.424)
Observations	116,350	2,166	1,523
LL	-152,915.300	-3,522.389	-2,688.419
AIC	305,916.600	7,166.777	5,498.839
BIC	306,332.200	7,513.296	5,823.874

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Appendix B

There is no reason to expect that the results in either the homophily or downstream models should hold only after conditioning on covariates. In this Appendix, I present results from both the main and downstream models without controls. The results, found in Table S3 and S4, are unchanged, with or without controls.

Table S3: Homophily models, with and without controls

	Proportion of Friends who Smoke					
	Contextual Effects		Contextual Effects x PGS		Contextual Effects x PGS (Whites)	
	w/o Controls	w/Controls	w/o Controls	w/Controls	w/o Controls	w/Controls
Daily Smoker x Smoking Punishment (First Offense)	0.019*	0.021*	0.045**	0.050**	0.043	0.057*
	(0.009)	(0.010)	(0.014)	(0.015)	(0.027)	(0.029)
Daily Smoker x Smoking Punishment (First Offense) x Smoking PGS			-0.032*	-0.039*	-0.031	-0.048
			(0.015)	(0.017)	(0.032)	(0.034)
Daily Smoker x Smoking PGS			0.076***	0.081***	0.015	0.046
			(0.015)	(0.016)	(0.033)	(0.034)
Smoking Punishment (First Offense) x Smoking PGS			0.006	-0.005	-0.005	-0.005
			(0.013)	(0.015)	(0.035)	(0.036)
Daily Smoker	0.408***	0.294***	0.362***	0.267***	0.409***	0.275***
	(0.009)	(0.010)	(0.013)	(0.016)	(0.027)	(0.029)
Smoking Punishment (First Offense)	0.005	-0.011	0.035**	0.014	0.045†	0.003
	(0.009)	(0.016)	(0.013)	(0.020)	(0.027)	(0.035)
Smoking PGS			0.152***	0.033	0.073*	0.025
			(0.013)	(0.028)	(0.035)	(0.039)
Constant	0.018*	-0.084	0.027*	-0.163†	0.116***	-0.116
	(0.009)	(0.068)	(0.013)	(0.085)	(0.028)	(0.139)
Observations	11,412	8,704	5,535	4,387	3,339	2,776
LL	-15,220.17	-10,885.89	-7,324.27	-5,541.61	-4,823.61	-3,735.03
AIC	30,448.35	21,853.79	14,664.54	11,191.22	9,663.22	7,574.07

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table S4: Downstream models, with and without controls

	Smoking Behavior (W4)			
	All Respondents	All Respondents	Smokers Only	Smokers Only
Proportion of Friends who Smoke x Smoking Punishment (First Offense)	0.446***	0.283*	0.425**	0.307†
	(0.124)	(0.142)	(0.157)	(0.184)
Proportion of Friends who Smoke	2.967***	0.641***	2.445***	0.565**
	(0.129)	(0.159)	(0.162)	(0.200)
Smoking Punishment (First Offense)	0.321*	-0.089	0.297	-0.209
	(0.127)	(0.224)	(0.190)	(0.308)
Constant	7.280***	25.622***	9.678***	29.237***

	(0.127)	(1.538)	(0.189)	(2.168)
Observations	8,648	6,495	5,005	3,778
LL	-33,648.790	-24,678.760	-19,990.180	-14,767.000
AIC	67,305.590	49,445.530	39,988.360	29,622.010

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Appendix C

In this Appendix, I examine the 12th grade results more closely. Implied in the text is that 12th graders are punished less, and this may attenuate the effect of smoking punishment on smoking homophily. In Table S5, I show that 12th graders are indeed punished less. In regressions, available upon request, I found that this pattern holds, and even grows more apparent, after controlling for age, isolating the effect of grade. This is also not because 12th graders smoke less than younger students as found in Table S6.

Table S5: Proportion of students in each high school grade who have and have not been suspended in the past year

		Grade			
		9	10	11	12
Suspended in past year	0	0.69	0.70	0.71	0.74
	1	0.31	0.30	0.29	0.26

Table S6: Proportion of students in each high school grade who smoke

Grade	Proportion Daily Smoker
9	0.22
10	0.22
11	0.24
12	0.25

Appendix D

Conceptually, the study design in this paper is akin to intent-to-treat. The main treatment variable, smoking punishment, is drawn from information reported by each school about their policy on how smoking is punished and therefore cannot speak to whether students are *actually* punished. The threat of punishment, as I find, appears to matter; however, it may also be worth examining whether students are indeed punished and if experiencing punishment or being in an environment in which punishments are carried out produces a similar effect. In Table S7, I use the number of smokers ever suspended in a high school as a measure for how strict a punishment environment is in place of the stated policy. In addition to substituting the measure of punishment environment, I also add two controls compared to the tables in the main text. The first is whether the respondent has ever been suspended, which I include to account for the respondents' potential contribution to the school-level measure. The second is the number of students who have been suspended in general to adjust for the fact smokers will be more likely to be suspended in schools where more students are suspended, independent of the risk that they take on by being smokers. Overall, the results look similar to those in the main text, except that the PGS result here is significant for whites but not when students of all races are included. However, while the results are compelling, I present the smoking policy results in the main text for the important reason that while we know who has been suspended, we have no information as to why they were suspended. I therefore cannot distinguish if the suspended smokers were punished for smoking or some other behavior.

Table S7: Using Number of Smokers Ever Suspended in High School as a measure of whether smokers are actually punished in a school environment

	Individual -level	School- level	Proportion of Friends who Smoke		
			Multilevel	Contextual Effects	Contextual Effects x PGS Contextual Effects x PGS (Whites)
Daily Smoker x Number of Smokers Ever Suspended in High School				0.001*** (0.0002)	0.002*** (0.0004) 0.003*** (0.001)
Daily Smoker x Number of Smokers Ever Suspended in High School x Smoking PGS					0.096*** (0.020) 0.093* (0.040)
Daily Smoker x Smoking PGS					0.0004 (0.001) 0.001 (0.001)
Number of Smokers Ever Suspended in High School x Smoking PGS					-0.001 (0.001) -0.002* (0.001)
Daily Smoker	0.283*** (0.010)		0.278*** (0.010)	0.245*** (0.012)	0.210*** (0.019) 0.184*** (0.034)
Ever suspended in high school	0.286*** (0.028)		0.285*** (0.028)	0.281*** (0.028)	0.258*** (0.041) 0.398*** (0.061)
Age	0.091*** (0.010)		0.091*** (0.011)	0.093*** (0.011)	0.104*** (0.015) 0.162*** (0.020)
Female	0.034*** (0.009)		0.036*** (0.009)	0.036*** (0.009)	0.039** (0.013) 0.065*** (0.018)
Black	-0.164*** (0.012)		-0.164*** (0.013)	-0.166*** (0.013)	-0.007 (0.052)
Hispanic	-0.039** (0.012)		-0.025† (0.013)	-0.026* (0.013)	0.034 (0.027)
GPA	-0.077*** (0.011)		-0.077*** (0.011)	-0.075*** (0.011)	-0.077*** (0.015) -0.083*** (0.021)
PVT	-0.019† (0.011)		-0.022† (0.012)	-0.023* (0.012)	-0.021 (0.017) -0.072** (0.025)
Married Household	-0.034** (0.013)		-0.035** (0.013)	-0.035** (0.013)	-0.026 (0.019) -0.043 (0.028)
Maternal Education	-0.008 (0.010)		-0.004 (0.011)	-0.004 (0.011)	-0.008 (0.015) -0.014 (0.022)
US Native	0.025* (0.010)		0.021† (0.011)	0.022† (0.011)	0.042* (0.015) 0.040 (0.022)

	(0.011)	(0.011)	(0.011)	(0.020)	(0.051)
Single Parent Household	0.008	0.007	0.006	0.012	0.018
	(0.012)	(0.012)	(0.012)	(0.018)	(0.028)
Family Income	0.004	0.006	0.006	-0.004	0.002
	(0.010)	(0.010)	(0.010)	(0.015)	(0.021)
Household Smoker	0.060***	0.056***	0.057***	0.051***	0.060***
	(0.009)	(0.009)	(0.009)	(0.013)	(0.017)
Network Centrality	-0.032**	-0.038***	-0.038***	-0.056***	-0.064***
	(0.010)	(0.010)	(0.010)	(0.014)	(0.019)
Has Older Siblings	0.002	0.002	0.0003	0.008	0.014
	(0.010)	(0.010)	(0.010)	(0.015)	(0.022)
All PC 1				0.164**	
				(0.055)	
All PC 2				0.001	
				(0.023)	
All PC 3				0.056*	
				(0.023)	
All PC 4				0.006	
				(0.016)	
White PC 1					0.329**
					(0.113)
White PC 2					0.052
					(0.079)
White PC 3					0.002
					(0.025)
White PC 4					0.014
					(0.074)
Number of Smokers Ever Suspended in High School	0.003*	0.003**	0.003**	0.002	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Smoking PGS				0.018	-0.008
				(0.031)	(0.045)
Mean Age	-0.010	-0.041†	-0.042*	-0.072**	-0.049
	(0.021)	(0.021)	(0.021)	(0.027)	(0.036)
Proportion Female	0.012	0.020†	0.019	0.019	-0.008
	(0.013)	(0.012)	(0.012)	(0.016)	(0.021)
Proportion Black	0.036	0.121**	0.122**	0.011	-0.058
	(0.040)	(0.038)	(0.038)	(0.152)	(0.243)
Proportion Hispanic	-0.030	-0.021	-0.022	-0.069	-0.097
	(0.041)	(0.040)	(0.040)	(0.061)	(0.092)
Mean GPA	0.017	0.031	0.032	0.019	-0.003

	(0.022)	(0.021)	(0.021)	(0.026)	(0.036)
Mean PVT	0.061	0.034	0.033	0.137*	0.038
	(0.040)	(0.038)	(0.038)	(0.061)	(0.086)
Proportion Married Household	-0.007	-0.017	-0.017	-0.010	-0.076
	(0.045)	(0.042)	(0.042)	(0.055)	(0.073)
Mean Maternal Education	-0.027	-0.006	-0.006	-0.022	-0.017
	(0.033)	(0.031)	(0.031)	(0.042)	(0.063)
Mean Family Income	-0.002	-0.005	-0.005	-0.024	0.00002
	(0.027)	(0.026)	(0.026)	(0.031)	(0.043)
Proportion w. Household Smoker	0.025	0.013	0.012	-0.016	-0.011
	(0.028)	(0.026)	(0.026)	(0.034)	(0.046)
Mean Network Centrality	0.024	0.036†	0.036†	0.052†	0.049
	(0.021)	(0.020)	(0.020)	(0.026)	(0.039)
Proportion w. Older Siblings	0.028	0.023	0.023	0.018	-0.010
	(0.022)	(0.021)	(0.021)	(0.027)	(0.037)
Proportion Daily Smoker	0.270***	0.197***	0.204***	0.243***	0.166**
	(0.033)	(0.031)	(0.031)	(0.038)	(0.053)
Proportion US Native	-0.042	-0.044	-0.044	-0.078†	0.065
	(0.029)	(0.027)	(0.027)	(0.046)	(0.068)
Proportion in Single Parent Household	0.036	0.0004	0.0003	0.024	0.003
	(0.043)	(0.041)	(0.041)	(0.050)	(0.066)
Punishment Profile	-0.011	-0.013	-0.012	-0.025	-0.025
	(0.015)	(0.014)	(0.014)	(0.017)	(0.022)
Number of Students Ever Suspended in High School	-0.013	-0.054*	-0.051*	-0.042	-0.046
	(0.025)	(0.023)	(0.023)	(0.029)	(0.039)
Mean Smoking PGS				0.290	-0.046
				(0.208)	(0.256)
Mean All PC 1				-0.411	
				(0.264)	
Mean All PC 2				-0.133†	
				(0.076)	
Mean All PC 3				-0.039	
				(0.031)	
Mean All PC 4				0.117*	
				(0.047)	
Mean White PC 1					-0.268
					(0.376)
Mean White PC 2					0.126
					(0.119)

Mean White PC 3						0.175† (0.092)
Mean White PC 4						0.023 (0.051)
Number of Classmates in In-home Sample		-0.077 (0.049)	-0.096* (0.044)	-0.076† (0.044)	-0.139† (0.075)	0.074 (0.117)
School Size		0.002 (0.023)	0.018 (0.022)	0.018 (0.022)	0.056† (0.030)	0.054 (0.040)
Region: Midwest		-0.058 (0.052)	-0.020 (0.047)	-0.021 (0.047)	-0.008 (0.062)	-0.045 (0.079)
Region: South		-0.023 (0.051)	0.003 (0.048)	0.001 (0.048)	0.029 (0.066)	-0.002 (0.093)
Region: Northeast		-0.072 (0.054)	-0.053 (0.051)	-0.053 (0.051)	-0.045 (0.074)	-0.005 (0.092)
Constant	-0.003 (0.023)	-0.042 (0.069)	-0.141* (0.065)	-0.129* (0.065)	-0.212* (0.089)	-0.179 (0.141)
Observations	8,761 -	11,521 -	8,744	8,744	4,413	2,801
LL	11,012.18 0	15,817.3 80	- 10,881.900	- 10,870.740	-5,540.763	-3,731.472
AIC	22,062.36 0	31,686.7 60	21,847.810	21,827.470	11,193.530	7,570.943
BIC	22,196.84 0	31,877.9 10	22,145.000	22,131.750	11,551.500	7,891.581

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Appendix E

In this Appendix, I examine whether there are any systematic differences between schools that punish harshly and those that do not. To do so, I run simple balancing tests evaluating whether mean demographics in schools differ systematically between levels of the smoking punishment variable. According to ANOVAs, run for each demographic and context-level variable, differences between schools are insignificant. The results can be found in Table S8.

Table S8: Balancing Tests for Smoking Punishment Harshness

Smoking Punishment Harshness (First Offense)	Mean Age	Prop. Female	Prop. Black	Prop. Hispanic	Mean Family Income	Mean Maternal Education	Prop. Household Smoker	State-level Excise Tax
Verbal Warning	16.74	0.50	0.19	0.14	3.63	5.68	0.45	29.40
Minor Action	16.48	0.51	0.20	0.10	3.57	5.79	0.46	42.77
In-school Suspension	16.56	0.51	0.17	0.18	3.51	5.19	0.45	32.95
Out-of-school Suspension	16.22	0.50	0.24	0.09	3.50	5.32	0.49	29.35
Significant?	No	No	No	No	No	Marginally	No	Marginally

Supplementary Information for Chapter 4

Appendix A: List of behaviors and their categorization

Risky Behaviors

Drinking: During the past twelve months, how often did you drink beer, wine, or liquor?

Smoking: During the past twelve months, how often did you smoke cigarettes?

Weed: Have you ever tried marijuana?

Hard Drugs: Have you ever tried hard drugs? (Summary measure across cocaine, glue or solvents (inhalents), LSD, PCP, ecstasy, mushrooms, speed, ice, heroin, or pills, without a doctor's prescription)

Fighting: In the past 12 months, how often did you get in a serious physical fight?

Graffiti: In the past 12 months, how often did you paint graffiti or signs on someone else's property or in a public place?

Motorcycle: During the past 12 months, how often did you ride a motorcycle?

Absences: During this school year, how many times {HAVE YOU SKIPPED/DID YOU SKIP} school for a full day without an excuse?

Sex and Relationships

Had Sex: Have you ever had sexual intercourse?

Number of Sex Partners per Year: Since January 1, 1994, with how many people in total have you had a sexual relationship?

Romantic Relationship: In the last 18 months—since {MONTH, YEAR}—have you had a

special romantic relationship with any one?

Hang out with Friends: During the past week, how many times did you just hang out with friends?

Entertainment

Television: How many hours a week do you watch television?

Video Games: How many hours a week do you play video or computer games?

Home and Out-of-School Activities

Exercise: During the past week, how many times did you do exercise, such as jogging, walking, karate, jumping rope, gymnastics or dancing?

Housework: During the past week, how many times did you do work around the house, such as cleaning, cooking, laundry, yardwork, or caring for a pet?

Arts: During the past week, how many times did you do hobbies, such as collecting baseball cards, playing a musical instrument, reading, or doing arts and crafts?

Skating and Biking: During the past week, how many times did you go roller-blading, roller-skating, skate-boarding, or bicycling?

Attend Religious: In the past 12 months, how often did you attend religious services?

Work: In the last 4 weeks, did you work—for pay—for anyone outside your home? This includes both regular jobs and things like baby-sitting or yard work.

Excluding GPA, behaviors in the following three categories were all drawn from the question:

Here is a list of clubs, organizations, and teams found at many schools. Darken the oval next to any of them that you are participating in this year, or that you plan to participate in later in the school year.

Big 3 Sports

Baseball

Basketball

American Football

Other Team Sports

Cheer

Soccer

Track

Volleyball

School Clubs and Academics

Honor Society

Student Life (participated in at least one of: Newspaper, Student Council and Yearbook)

Language Club (participated in at least one of: French, German, Latin, or Spanish Club)

Subject Club (participated in at least one of: Book Club, Computer Club, Debate Team, Drama Club, History Club, Math Club, Science Club)

Music Club (participated in at least one of: Band, Orchestra, Choir)

GPA (mean of their GPA in English, Math, Social Science, and Science classes)