Leveraging variation in treatment effects to understand mechanisms for universal, preventative interventions: A framework, empirical examples, and a new statistical tool.
Chair: David S. Yeager

This symposium will frame research into variation in treatment effects—also called "moderators," "treatment effect heterogeneity," "subgroup effects," or "effect modification"—as an opportunity to build interdisciplinary theory about the mechanisms that allow time-limited and scalable interventions to show enduring effects on social, emotional, or educational outcomes over time. The ultimate goal of the symposium is to reduce emphasis on average treatment effects and overly-simplistic judgments of "replication" and "failure to replicate," and instead show how—both in the abstract and in high-profile empirical case studies—principled interrogations of where intervention effects endure and where they do not can reveal new theoretical insights into mechanism. A secondary goal is to provide concrete examples and statistical tools for how analysts can design new studies to learn from effect heterogeneity.

After the chair (Yeager) provides an overview of the subject area, the first speaker (Yamamoto) will present a unifying framework that connects research on *mediators* as a means for understanding mechanisms with approaches that examine *moderators*. This talk will ground the subsequent talks in the Rubin counterfactual model and provide a common language for interpreting their results. Then, three empirical talks (Gopalan, Walton, and Bryan) will show data revealing moderation of treatment effects in school settings. These three cases are illustrative because they used rigorous methods—large-sample, double-blind, randomized designs and pre-registered analysis plans—but nevertheless discovered moderators that led to a deeper understanding of why a psychological intervention's effects endured. Finally, a talk from a team of Bayesian statisticians (Murray, Carvalho, and Hahn) will present a new, hierarchical, Bayesian, machine-learning method for understanding variation in treatment effects, and illustrate how audience members can access the R package and use it for their own analyses.

Papers:

Causal Mechanisms and Effect Modification
Teppei Yamamoto, MIT

Causal mediation and moderation are often contrasted as two distinct concepts for describing the role of a third variable in a cause-effect relationship. Although both concepts are broadly considered to be related to the notion of mechanisms, they are typically analyzed via different causal models -- potential outcomes and sufficient component causes -- in the causal inference literature. In this paper, we provide a unified perspective where both mediation and moderation are defined as evidence of the same underlying causal mechanism. In doing so, we introduce the concept of "switch," or the effect modifier that has no effect on any mediator of the treatment effect except a particular mediator that represents the mechanism of interest. We show that the existence of such a switch is sufficient for the nonparametric identification of the average natural indirect effect. Our proposed framework also resolves an important limitation of Pearl's causal direct acyclic graph (DAG) framework in representing interactive relationships.

A Growth Mindset Intervention Designed to Improve High School GPA is Most Effective for Under-performing Students Attending Schools with Supportive Peer Norms
Maithreyi Gopalan, University of Pittsburgh, David S. Yeager, University of Texas at Austin, and the *National Study of Learning Mindsets* team

This talk will present the main moderation results from the *National Study of Learning Mindsets (NSLM)*. The study delivered a short *growth mindset* intervention—an intervention teaching that intellectual abilities are not fixed but can be developed—with the goal of understanding where the intervention redirected the educational trajectories of lower-achieving students. Three main findings emerged. First, this short, universal, preventative psychological intervention had modest but consequential effects on outcomes such as grades in core classes over the year (the primary outcome) and rates of taking advanced math courses the next year (an exploratory outcome). Second, the study identified a school-level factor—behavioral norms regarding challenging schoolwork—that moderated treatment effects on the primary outcome of grades; effects were stronger when the peer norms aligned with its message. Third, because the study employed a number of techniques to reduce false discoveries, such as independent data collection, pre-registration of analysis, and re-analysis of data by statisticians who fit a flexible and conservative Bayesian model, it provided an example for how to examine treatment effect heterogeneity in a way that is both reproducible and generalizable. This example could be followed in trials conducted in medicine, public health, policy analysis, and other allied disciplines, to better understand when and where interventions improve social well-being.

How Do the Effects of a Brief Social Belonging Intervention on College Achievement Vary Across Students and Schools?
Gregory Walton, Stanford University, and the College Transition Collaborative

This talk will present the main moderation findings from the College Transition Collaborative (CTC)'s trial of the social-belonging intervention, brief online reading-and-reflection module prior to college matriculation that represents common challenges to belonging as normal in the transition to college and as improving with time. Past trials at single institutional sites find that this exercise can improve college persistence among students from groups that are often disadvantaged in higher education, including negatively stereotyped racial-minority students and first-generation college students. CTC tested the belonging intervention in a multi-site, individual-random-assignment experiment in over 20 colleges and universities ($N > 45{,}000$). It found that disadvantaged students benefitted more from the intervention, provided that their identity group on campus was able to attain a moderate level of belonging by the spring of their first year in college, absent the treatment. That is, the social-belonging intervention was most effective for underperforming groups for whom the intervention's message could become "true" over time in their college setting.

A Brief "Values Alignment" Intervention Improves Dietary Preferences More Strongly Among Adolescents Higher in Testosterone
Christopher Bryan, University of Chicago Booth School of Business, Fortunato (Nick) Medrano and Robert Josephs, University of Texas at Austin

Prior research has found that 8th grade adolescents reduced their junk food consumption when they received a "values-aligned" treatment –one that framed healthy eating as a way to stand up against food companies who use marketing to control young people's choices. The supposed mechanism for this effect is an alignment between healthier eating and the adolescent values of autonomy from adult control and pursuit of social justice. This values-aligned framing supposedly increases the perception that healthier eating will be socially rewarding, and these perceived social rewards in turn may act as a self-sustaining cause of enduring behavior change. For the first time, this hypothesis is tested by examining whether long-lag behavioral treatment effects are moderated by pre-intervention levels of testosterone. Testosterone rises dramatically in males and females during pubertal maturation and has, in past human and animal studies conducted in laboratory settings, shown an effect of increasing the brain's responsiveness to social rewards. The pre-registered, randomized, field experiment finds that adolescents assigned to the values-aligned framing made healthier purchases in the lunchroom for three months after receiving the treatment, especially so if they were higher in testosterone. This finding reveals mechanisms for a brief intervention and also yields new insights into the functioning of a hormone that is key to pubertal maturation.

Using Bayesian Causal Forest Models to Examine Treatment Effect Heterogeneity
Jared Murray, Carlos Carvalho, Richard Hahn

This talk will introduce Bayesian causal forests (BCF), a flexible analytic method for estimating and interrogating effect heterogeneity, and illustrate how it has been used in past research (see: https://arxiv.org/abs/1706.09523). While flexible regression and machine learning methods have made dramatic advances in prediction contexts, estimating heterogeneous treatment effects with noisy multilevel data requires a more nuanced approach than simply applying off-the-shelf machine learning tools. This has been an active area of research in statistics and machine learning, and several empirical evaluations suggest that BCF is a top-performing method. We demonstrate the use of BCF for estimating and summarizing heterogeneous treatment effects, including identifying interesting subgroups. We discuss how BCF could enhance the analyses in the three case studies presented before it.