Title
Heavy Episodic Drinking and Academic Consequences in Adolescence: The Mediational Role of Neurocognitive Functioning

Permalink
https://escholarship.org/uc/item/5m21n1s4

Author
Cummins, Kevin

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Heavy Episodic Drinking and Academic Consequences in Adolescence: The Mediational Role of Neuropsychological Functioning

A dissertation submitted in partial satisfaction of the requirement for the degree Doctor in Philosophy

in

Interdisciplinary Research on Substance Use

by

Kevin Matthew Cummins

Committee in Charge:

University of California San Diego
Professor Sandra A. Brown, Chair
Professor Eileen V. Pitpitan
Professor Susan F. Tapert
Professor Ryan S. Trim

San Diego State University
Professor María Luisa Zúñiga, Co-Chair
Professor Mark Reed

2020

The Dissertation of Kevin Matthew Cummins is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

San Diego State University

2020

DEDICATION

This work is dedicated to my wish. I had no idea what I really should have wished for until we met. There is no way I could have completed my doctoral training without your initial strong, but sober, endorsement and continued patience. Very few partners would have comfortably prorogued birthday and anniversary celebrations for so many years. Thank you for making it possible for me to play with my friends in the science buildings for the last five years.

EPIGRAPH

There is nothing as practical as a good theory.

–*Kurt Lewin* (McCain, 2015, p. 1)

Field theory can hardly be called a theory in the usual sense.

–*Kurt Lewin* (Lewin, 1951, p. 45)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It has been among my highest honors and privileges to have had the mentorship and support of Sandra Brown. I am grateful for her encouragement, advice, and trust. Her mentorship has helped me advance, broaden, and integrate my knowledge and skills. I am deeply proud to be a member of the Brown Lab.

I must also make special mention of Susan Tapert. She has welcomed me into her lab and provided me with critical nudges and insight that helped shape the course of my progress. Through their uncommon dedication, both she and Sandra Brown have made it possible for me to complete this dissertation.

The whole NCANDA team has been generous with their support. Sonja Eberson, Alyssa Lopez, Matthea Monroe, Laura Fiero, Ty Brumbach, and Natasha Wade deserve a special thanks for their efforts to accommodate and support me.

I also must recognize Gail Laughlin and Robert Riffenburgh for encouraging me to return as a student to seek a doctorate degree. I also must acknowledge Boyd Collier for making me believe that a high *g* is not as important as curiosity, perseverance, stamina, and knowing how to capitalize on mistakes, which is a perspective I take into all my classrooms regardless of if I am a student or instructor. Dr. Collier also encouraged me to apply for my ARCS scholarship that funded my prior graduate training, which set everything in motion. Dawn Lawson generously donated her knowledge of ecology in discussions about early versions of Chapter 2. She has no idea but she has been a role model for me since 1995, where she was the coolest and most courageous graduate students in the department. She always spoke her mind without being outspoken, and had everyone's attention and humbly absorbed everyone's input.

My cohort was also instrumental in shaping the way I think and work. Mario Morales, Sarah Clingan, and Nicole Pepper have shared their unique histories, knowledge, and disciplinary perspectives with me throughout my training. It has been my privilege to train with them.

I have been calling my committee the dream team because of their varied and complementary skills, shared research interests, and extremely high regard that students have for each committee member as researchers, mentors, colleagues, and friends.

Three will never be a program director who gives more of themselves to the program and its indivdiual students than Dr. Zúñiga. Dr. Pitpitan, I am so grateful for your support in filling in a gap in my analytical toolbox and helping me to navigate through challenging findings. Dr. Reed, your generosity with your time and interest and support for advancing some of my most unconventional ideas may have helped sparked one of the lines of inquiry that I run with during my career.

The dissertation would not be possible without the financial support I received. Firstly, I would like to thank San Diego State University for tuition and stipend support. Additionally, I would like to acknowledge support from the NIAAA grant number U24 AA021695-08.

Finally, I would like to thank the participants who devoted their time and energy to this project.

Chapter 2 has been submitted for publication and is under review as: Cummins, K. M., Zuniga, M. L., Reed, M., Pitpitan, E. V., & Brown, S. A. Reforming the use of theory in health promotion: lessons from ecology. *Health Psychology Reviews*. Kevin Cummins was the primary investigator and lead author of this paper. This paper was co-authored by María Luisa Zúñiga, Mark Reed, Sandra Brown and Eileen Pitpitan.

Chapter 3 has been submitted for publication and is under review: Cummins, K. M., Pitpitan, E. V., Brown, S. A., Clark, D. B., Brumback, T., Trim, R. S., & Tapert, S. F. Comparison of factor analysis models applied to the NCANDA neuropsychological test battery. *Journal of Clinical and Experimental Neuropsychology*. Kevin Cummins was the primary investigator and author of this paper.

Chapter 4 is currently being prepared for submission for publication as: Cummins, K. M., Brown, S. A., Trim, R. S., & Tapert, S. F. Mediation of heavy episodic drinking effects on academic performance through neurocognitive functioning. Kevin Cummins was the primary investigator and author of this paper.

VITA

2020	Doctor of Philosophy in Interdisciplinary Research on Substance Use University of California San Diego and San Diego State University (San Diego, CA)
2005	Master of Science (Statistics) San Diego State University (San Diego, CA)
2003	Master of Science (Biology) San Diego State University (San Diego, CA)
1995	Bachelor of Science (Biology) San Diego State University (San Diego, CA)

SELECTED PUBLICATIONS

1. Bekman, N. M., **Cummins, K.**, & Brown, S. A. (2011). The influence of alcohol-related cognitions on personality-based risk for alcohol use during adolescence. *Journal of Child & Adolescent Substance Abuse*, 20(4), 387-405.

2. Black, C. H., **Cummins, K. M.**, Lawson, D. M., & Cobb, C. (2016). Using wildfires as a natural experiment to evaluate the effect of fire on southern California vernal pool plant communities. *Global Ecology and Conservation*, 7, 97-106.

3. Bravo, M. M., **Cummins, K.**, Nessler, J. A., & Newcomer, S. C. (2015). Heart rate responses of high school students participating in surfing physical education. *Journal of Strength and Conditioning Research*, 30(2), 1721-1726.

4. Brown, S. A., Brumback, T., Tomlinson, K., **Cummins, K**., Thompson, W. K., Nagel, B. J., . . . Chung, T. (2015). The National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA): a multisite study of adolescent development and substance use. *Journal of Studies on Alcohol and Drugs*, 76(6), 895-908.

5. Cui, R., Tate, S. R., **Cummins, K.**, Skidmore, J. R., & Brown, S. A. (2015). Chronic physical health problems moderate changes in depression and substance use among dual diagnosed individuals during and after treatment. *Substance Use and Misuse*, 50(2), 174-183.

6. **Cummins, K.** (1999). To repeat or block in experiments? *Trends in Plant Sciences*, 7(43), 251-252.

7. **Cummins, K**., Diep, S., & Brown, S. (2019). Alcohol expectancies moderate the association between school connectedness and alcohol consumption. *Journal of School Health*, 89(11), 865-873.

8. **Cummins, K.** (2009). Tips on writing results for a scientific paper. *AmStat News*, September, 39-41.

9. **Cummins, K. M.,** & Marks, C. L. (2020). Farewell to Bright-Line: A Guide to Reporting Quantitative Results Without the S-Word. *Frontiers in Psychology*, Accepted.

10. Haller, M., Norman, S. B., **Cummins, K**., Trim, R. S., Xu, X., Cui, R., . . . Tate, S. R. (2016). Integrated cognitive behavioral therapy versus cognitive processing therapy for adults with depression, substance use disorder, and trauma. *Journal of Substance Abuse Treatment*, 62, 38-48.

11. Hanson, K. L., **Cummins, K.**, Tapert, S. F., & Brown, S. A. (2011). Changes in neuropsychological functioning over 10 years following adolescent substance abuse treatment. *Psychology of Addictive Behaviors*, 25(1), 127-134.

12. Jassal, S. K., Chonchol, M., Laughlin, G. A., **Cummins, K. M**., Smits, G., Kramer, C. K., . . . Barrett-Connor, E. (2012). Kidney function and progression of coronary artery calcium in community-dwelling older adults (from the Rancho Bernardo Study). *The American Journal of Cardiology*, 110(10), 1425-1433.

13. Larsen, B. A., Laughlin, G. A., & **Cummins, K**. (2017). Adipokines and severity and progression of coronary artery calcium: Findings from the Rancho Bernardo Study. *Atherosclerosis*, 1, 1-6.

14. Laughlin, G. A., **Cummins, K. M.,** Wassel, C. L., Daniels, L. B., & Ix, J. H. (2012). The association of fetuin-A with cardiovascular disease mortality in older community-dwelling adults: the Rancho Bernardo study. J*ournal of the American College of Cardiology*, 59(19), 1688-1696.

15. Laughlin, G. A., Ix, J. H., **Cummins, K.,** Allison, M. A., & Daniels, L. B. (2012). Extremes of an aromatase index predict increased 25-year risk of cardiovascular mortality in older women. *Clinical Endocrinology*, 77(3), 391-398.

16. Lydecker, K. P., Tate, S. R., **Cummins, K. M.,** McQuaid, J., Granholm, E., & Brown, S. A. (2010). Clinical outcomes of an integrated treatment for depression and substance use disorders. *Psychology of Addictive Behaviors*, 24(3), 453.

17. Pfefferbaum, A., Kwon, D., Brumback, T., Thompson, W. K., **Cummins, K**., Tapert, S. F., . . . Prouty, D. (2017). Altered brain developmental trajectories in adolescents after initiating drinking. *American Journal of Psychiatry*

18. Pulido, C., Brown, S. A., **Cummins, K**., Paulus, M. P., & Tapert, S. F. (2010). Alcohol cue reactivity task development. *Addictive Behaviors*, 35(2), 84-90.

19. Riffenburgh, R. H., & **Cummins, K. M**. (2006). A simple and general change point identifier. *Statistics in Medicine*, 25(6), 1067-1077.

20. Tomlinson, K. L., **Cummins, K. M.,** & Brown, S. A. (2013). Social anxiety and onset of drinking in early adolescence. *Journal of Child & Adolescent Substance Abuse*, 22(2), 163-177.

ABSTRACT OF THE DISSERTATION

Heavy Episodic Drinking and Academic Consequences in Adolescence:
The Mediational Role of Neuropsychological Functioning

by

Kevin Matthew Cummins

Doctor in Philosophy in Interdisciplinary Research on Substance Use

University of California San Diego, 2020
San Diego State University, 2020

Professor Sandra A. Brown, Chair
Professor María Luisa Zúñiga, Co-Chair

Background: A growing confluence of evidence demonstrates that heavy episodic drinking (HED) during adolescence can alter neurodevelopment. Alcohol use has a well-established and strong association with academic functioning. The extent to which this association is mediated by neuropsychological decrements remains a critical gap in the research literature.

Aims: Aim 1: Articulate an adaptable conceptual and procedural approach to guide research on the joint study of HED, neuropsychological functioning, and academic performance. Aim 2: Examine the utility of applying factor analysis frameworks to the measurement of neuropsychological functioning to optimize the evaluation of HED effects. Aim 3: Evaluate the potential mediating role of neuropsychological functioning in the association between adolescent alcohol use and academic performance.

Methods: Aim 1 was addressed with a critical review of the literature used by public health, social sciences, behavioral sciences, biological sciences, and philosophy to guide the use of theory in scientific investigations. Measurement models for the neuropsychological assessment used in this project were established through comparative confirmatory factor modeling as a means to address Aim 2. Causal models based on propensity score weighting were used to assess the association between HED attributable decrements in neuropsychological function and changes in academic performance. Aims 2 and 3 were addressed with data obtained from the NCANDA study, which is a longitudinal study of youth (ages 14-18, n = 419).

Results: Challenges to current approaches to the use of theory were identified, and solutions, including paratheoretical framing and abductive inference, were outlined. The conceptually based Gur+ and 8-Factor models were found to be adequate representations of the factor structure and applicable to the investigation of the effects of HED on neuropsychological functioning. Prior reports of the association between neuropsychological functioning and grade point average were corroborated. Causal effects of HED mediated through neuropsychological changes on academic performance were not detected.

Conclusions: Although causal effects were not detected, the theoretical and analytical paradigms developed in this project are applicable to future investigations, including the parent study of this project as more NCANDA participants escalate their drinking exposures.

CHAPTER 1: INTRODUCTION

BACKGROUND

Alcohol is the most commonly used substance among high school students in the United States (Johnston et al., 2018). Sixty-two percent of high school seniors have consumed alcohol (Johnston et al., 2018). Almost three-quarters (73%) of seniors with a drinking history have been drunk (Johnston et al., 2018). Among secondary school students, 10% report HED in the prior two weeks (Miech et al., 2017). Approximately one fifth (~18%) of seniors had engaged in heavy episodic drinking (HED; 4/5+ drinks in a row) in the prior two weeks (Johnston et al., 2018). Although the prevalence of recent HED in the United States has dropped by 55% since its peak in 1997, the decline observed throughout the early 2000s has leveled off (Twenge & Park, 2019).

HED remains a public health concern. Globally, alcohol is the strongest risk factor for poorer disability-adjusted life-years among 15-24-year-olds (Gore et al., 2011). Alcohol is an attributable factor in 17% of deaths among 15-19-year-olds in North America (WHO, 2014). The rate is even higher in Europe (20%; WHO, 2014). In most developed western nations, adolescent alcohol involvement exceeds that experienced in the United States (Ahlström & Österberg, 2004). Additionally, culminating evidence implicates adolescence as a sensitive period for the potentiating effects of alcohol exposure to later risk of addiction, which is a major contributor to the global burden of disease (Rehm, Mathers, Popova, & Thavorncharoensap…, 2009; Rehm & Imtiaz, 2016; Rehm et al., 2017; Whiteford, Degenhardt, Rehm, & Baxter…, 2013). Alcohol use among youth presents various public health concerns.

ALCOHOL AS A NEUROTOXIC AGENT

Alcohol consumption can result in brain damage. Prolonged exposure to high levels of alcohol, as experienced by some adults with alcohol use disorder, can result in brain atrophy as a consequence of behaviorally mediated and direct neurotoxic mechanisms (Oscar-Berman & Marinkovic, 2003). Noted brain regions that are susceptible include areas involved in learning and memory (basal forebrain and limbic system), posture and motor control (cerebellum) and, particularly, executive functioning (prefrontal cortex)(Sullivan, Deshmukh, Desmond, Lim, & Pfefferbaum, 2000). White matter is broadly vulnerable to the chronic effects of alcohol exposure in humans (Zahr & Pfefferbaum, 2017). The neurocognitive consequences can result in memory, learning, inhibition and executive functioning deficits, in addition to emotional blunting. Profound disability can result (Oscar-Berman & Marinkovic, 2003). However, these findings were born from the study of adults with years of excessive alcohol exposure. This work my not translate to exposures experienced by community adolescents, where typical drinking patterns are intermittent and years of exposure limited.

Adolescence is an important period of neuromaturation and a sensitive period for persistent alterations of the brain’s reward circuitry to alcohol exposure in animal models (Crews, Vetreno, & Broadwater…, 2016; Salling et al., 2018) and early drinking onset is a known risk factor for alcohol use disorder in humans (Grant et al., 2006). There is also emerging evidence that patterns of alcohol exposure experienced by youth in the community can result in alterations in neuroanatomical development (Squeglia, Spadoni, Infante, Myers, & Tapert, 2009b). The functional implications of alcohol-induced neuromodulation are just now being assessed in large scale community samples (Brown et al., 2015; Jernigan, Brown, & Dowling, 2018; Akshoomoff et al., 2014).

Multiple lines of evidence converge showing that adolescence is a particularly vulnerable period for the neurological impacts of alcohol exposure. Cognitive deficits associated with adolescent alcohol exposure include verbal learning (Mahmood, Jacobus, Bava, Scarlett, & Tapert, 2010; Sneider, Cohen-Gilbert, Crowley, Paul, & Silveri, 2013), visuospatial processing (Tapert & Brown, 1999), executive functioning (Mota et al., 2013), attention (Tapert & Brown, 1999; Tapert, Granholm, Leedy, & Brown, 2002) and memory (Mota et al., 2013; Hanson, Cummins, Tapert, & Brown, 2011; Tapert et al., 2002). A recent longitudinal study of 112 initially abstinent community adolescents who began drinking prior to a six-year follow up found decrements in short-delay memory and verbal learning associated with exposure to more extreme alcohol exposures (Nguyen-Louie et al., 2016). Although the unexplained variance remained high, alcohol exposure (peak eBAC) was linearly related to most of the neuropsychology measures. The most extreme binge drinkers (the top 31% of the sample) recall 8-12% fewer words than the lowest intensity drinking class. In a smaller longitudinal analysis of 3-year outcomes, HED exposure prospectively predicted to poorer visuospatial functioning among girls and hangover events predicted to lower attention in boys, where the exposures explained up to 10% of the observed variation in neuropsychology measures (Squeglia et al., 2009b). These are exemplars of the current research base utilizing human subjects. Fundamental limitations for many of these observational studies are related to small sample sizes.

The substantial experimental rodent research enhances the support of causal inferences. These studies also allow for the disentanglement of the distinct effects of different substances. Alcohol exposure has been found to reduce cognitive flexibility, increase anxiety-like behavior, increase motivation for alcohol, and persistence of adolescent behavioral phenotypes into adulthood in animal models (Spear, 2018; Crews et al., 2016). However, animal models have limitations in their correspondence to human biology, human socialization, and the complexity of learned behavior.

Understanding the association between substance use, neurocognitive functioning, and academic functioning is also complicated by the knowledge that preexisting neurocognitive deficits predict features of youth substance use involvement (Squeglia & Gray, 2016; J. Conrod & Nikolaou, 2016; Spear, 2018). Much of the correlational research may reflect differences that predate substance use onset. Recently initiated large scale longitudinal community studies of demographically representative substance use naïve youth provide opportunities to solidify and extend the early findings.

NEUROCOGNITIVE AND ACADEMIC FUNCTIONING

Although neurocognitive functioning can be measured along varied dimensions, most measures that are conceptualized as components of traditional intelligence generally share a substantial amount of common variance in community samples (Kamphaus, 2005). The first principal component (g-factor) accounts for greater than 90% of the variance in batteries of cognitive functioning (Spearman, 1927). Prospective studies have found correlations ranging from 0.40 to 0.63 between baseline neurocognitive assessment scores and measures of educational attainment (Jencks, 1979). Grade point average and measures of neuropsychological functioning is similarly correlated (Roth et al., 2015; Mackintosh & Mackintosh, 2011). Most of these estimates are attenuated by unknown levels of measurement error. When cognitive ability was measured broadly with relatively high reliability, the estimated prospective correlation with standardized academic scores has been found to reach 0.81 (Deary, Strand, Smith, & Fernandes, 2007). Although concerns that many early cognitive tests strongly reflected academic achievement, there are measures that are relatively delineated from scholastic skills and experiences (Kamphaus, 2005). Further, longitudinal research suggests that cognitive functioning more strongly influences later academic success than academic performance influences subsequent measures of cognitive functioning (Watkins, Lei, & Canivez, 2007).

Academic performance and achievement are influenced by factors beyond those measured in typical assessments of scholastic aptitude. Many of these influences are also generated or mediated through neurocognitive processes. These include factors such as teacher pedagogy, parental involvement, resilience, motivation, and behavioral engagement. Other influences, such as a youth's socio-economic setting, are found at higher ecological levels. Many of these alternative influences may be dominant determinants of scholastic functioning.

There are also putative common prior causes of adolescent substance use involvement and academic functioning. Important shared antecedents include parental education and marital status (Cohen & Rice, 1997; Bachman et al., 2008). Primary school setbacks, including being held back, are also predictive of both greater involvement in substance use and weaker academic functioning in high school (Hawkins, Catalano, & Miller, 1992; Martin, 2011). Psychiatric symptomology is associated with both outcomes (Tomlinson, Cummins, & Brown, 2013; Kaplow, Curran, Angold, & Costello, 2001; Goldstein, 2009; Bachman et al., 2008; Giancola & Mezzich, 2000).

RESEARCH GAP

Numerous reciprocal and interacting factors predict students' academic success. Substance use is one of the factors that has robust relationships with academic performance and attainment (Cox, Zhang, & Johnson…, 2007; Lopez-Frias & Fernandez…, 2001; Singleton &

Wolfson, 2009; Friedman & Glickman…, 1985; Latvala et al., 2014; McCluskey, Krohn, & Lizotte…, 2002; Bergen, Martin, Roeger, & Allison, 2005; Engberg & Morral, 2006; Balsa, Giuliano, & French, 2011; Porter & Pryor, 2007; Bachman et al., 2008). High school dropouts are three times as likely to have a history of binge drinking in early adolescence when compared to youth who attain more than 3 years of college (Bachman et al., 2008). HED among adolescents is prospectively correlated with lower grade point average ($r > -0.15$)(Bachman et al., 2008). The association is strongest among adolescents with baseline disabilities. Grade point average was 0.54 points lower for learning disabled students who engaged in any binge drinking when compared to those who were abstinent (Hollar & Moore, 2004). Although common preexisting confounders (e.g., familial environment) may explain part of the association, the association is thought to be causal (Bachman et al., 2008). Some have postulated that adolescent substance use facilitation of the development of social networks with high levels of antisocial attitudes and behaviors is a key distal mechanism. The proximal mediator in this pathway is school disengagement. Early adolescent alcohol involvement is known to be negatively associated with lower academic success through multiple mechanisms (Bachman et al., 2008). The extent that neurocognitive decrements resulting from alcohol use compliment these pathways is not well established.

Conversely, low scholastic success can precipitate the escalation of alcohol use (Crum, Juon, Green, & Robertson…, 2006; Schulenberg, Bachman, O'Malley, & Johnston, 1994; Bryant, Schulenberg, & O'malley…, 2003). Regarding the causal mechanisms, Bachman et al. (2008) concluded that: 1) prior common causes are operating, 2) academic success modulates the risk of substance use, and 3) substance use impacts academic success. They postulate that the effect of substance use exposure, including alcohol, is the weakest causal link. However, this pathway has not been adequately quantified.

The literature has not disentangled and resolved the various mechanisms behind the effects of alcohol on academic functioning. A substantial body of work has accumulated in two areas: the association of neuropsychological functioning and academic success, and alcohol use and academic functioning. Additionally, a coalescing body of work is developing on the association between normative youth alcohol exposure and neuropsychological functioning. Although this work has progressed over the last decade, these three areas have not been sufficiently tied together, even though it has been over a decade since King, Meehan, Trim & Chassin (2006) noted the need to establish the relative importance of neuropsychological impairment attributable to alcohol exposure as a mechanism leading to reduced academic functioning.

The overarching aim of this dissertation was to address this research gap. The lack of resolution in the literature may be, in part, due to the challenge in estimating subtle effects in the context of complex and sometimes strong competing exposures and mechanisms. The consequential need for well-designed and high-powered studies has been a barrier for adequately estimating effects related to the research gap. Prior investigations have evaluated determinants in relative isolation. Work has not begun that quantitatively ties these determinants together. This gap is also made more challenging to address as there is no widely used framework or integrative literature that organizes and ranks these diverse relationships.

THEORETICAL CONTEXT AND AIMS

A pilot review conducted for this project of the recent literature (prior five years) that addresses the links in the HED to neurocognitive functioning to academic performance pathway found only 7 of 173 papers (4.0%) invoked a theory. Five papers explicitly used a framework. Two used conceptual models for contextualization. Examples include papers that invoke the Triadic Model, Dual Process Theory, and Incentive Motivation Model (Spada, Albery, & Moss, 2015; Weissman et al., 2015). None used theory as a source of quantified or risky novel predictions. None provided predictions or included mechanisms that would explain patterns directly relevant to the overarching aim of this dissertation project. None addressed a large number of the known determinants of academic performance or presented biological mechanisms that would allow for the prediction of the specific effects HED should have on various dimensions of neuropsychological functioning.

Numerous researchers have advanced the recommendation that theory be used to structure social science, and public health research in particular (Goodson, 2010; Rothman, 2004; Painter, Borba, Hynes, Mays, & Glanz, 2008; Glanz & Bishop, 2010). The putative benefits include advancement of the field’s knowledge and understanding, which can ultimately be used to improved intervention effectiveness. These recommendations are clouded by several features of the current theoretical landscape and ambiguity in the recommendations themselves.

This area of weak resolution is particularly critical to this study. It is unclear how to apply the recommendations when a study’s focus lies outside areas where an *established* theory is directly relevant. An established, or even tentative, theory for the phenomena under investigation have not been identified that can be used to place the study in a comprehensive context or make useful predictions. Because the current research is occurring in the context of a dearth of theoretical conceptualizations that address the domain of the study's overarching aim, a subservient aim (Aim 1) will provide an analysis of the consequences of proceeding without a conventional approach to theory. No established theory addresses the three-way association under investigation. Consequently, adapting tangentially related theory first requires selection and specification from among numerous competing, complementary, and amorphous theoretical schemes (e.g., framework, model) to organize, contextualize, and guide this research. The research and its outcomes could ultimately pivot on arbitrary or undisclosed subjective decisions. Completion of Aim 1 is expected to result in an outline of how science can advance in the absence of well-developed theory that can reliably guide researchers operating in such scenarios.

Aim 1: Articulate an adaptable conceptual and procedural approach to guide research on the joint study of HED, neuropsychological functioning, and academic performance.

To address the gap in the literature on the links between HED and academic performance data from the first four years of the National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) project were utilized. NCANDA's strength is its broad assessment of neurocognitive functioning.

The current literature and theoretical landscape provide little direct guidance on how to optimize and standardize the analyses of these data in the context of research about school and alcohol use. In the service of the overarching aim for this project, the optimum measurement model for neurocognitive constructs in NCANDA were established.

Aim 2: Examine the utility of applying factor analysis frameworks to the measurement of neuropsychological functioning to optimize the evaluation of HED effects.

The final aim directly addressed the important empirical questions that have been posed regarding the neurocognitive mediated impacts of HED on academic performance. Findings from Aim 1 were used to inform the structure and integration of the research approach used to address both Aim 2 and Aim 3. The results from Aim 2 were used to identify a measurement model used in the evaluation of Aim 3.

Aim 3: Evaluate the potential mediating role of neurocognitive functioning in the association between adolescent alcohol use and academic performance. *Hypothesis 3*: Neurocognitive performance weakly mediates a negative prospective association between HED and academic performance.

CHAPTER 2: REFORMING THE USE OF THEORY: LESSONS FROM ECOLOGY

ABSTRACT

Persuasive arguments for the use of theory have been influential in health promotion. The benefits of theory are expected to be substantial. The field has yet to broadly recognize that many of its conceptual devices that are treated as theory function more as models. This can explain why the evidence to support the use of theory in the development of interventions has not been conclusive. Health promotion is not alone in struggling to attain a strong network of theories. This is one way in which ecology parallels health promotion. However, ecology has taken a different route. Ecology's progress provides justification for expanding health promotion's lexicon to improve the delineation and characterization of its conceptual devices. It is posited that an improved lexicon will facilitate improved understanding of related methodological recommendations and underscore that suggested benefits are achieved through using well-formed, severely tested theories. Ecology's history suggests paratheoretical approaches can be successful in realms of inquiry that are resistant to theory formation such as health promotion. Research can be productive in the absence of theory through the use of abductive framing. Public health can enhance its theory-derived strengths by focusing its effort on modeling and developing general bounding principles.

INTRODUCTION

The development and application of theory holds a prominent role in health promotion (HP). The most sound, efficient, and effective intervention efforts are expected to be facilitated by theory utilization (Albada, Ausems, Bensing, & van Dulmen, 2009; Datta & Petticrew, 2013; Glanz & Bishop, 2010; Lippke & Ziegelmann, 2008; Noar, Benac, & Harris, 2007; Michie, Johnston, Francis, Hardeman, & Eccles, 2008). This justifies the increased valuation of work that includes an exposition of the project's theoretical basis. It has even been stated that leveraging theory is a necessary criterion for evidence-informed HP (Bartholomew Eldredge et al., 2016, p. 7) (DiClemente, Crosby, & Kegler, 2009, p. 10). However, there is an apparent acceptance of research that integrates theory as a post hoc façade module tacked onto manuscripts (Michie & Prestwich, 2010). If superficial use of theory is undetectable, or acceptable, it calls into question theory's ability to facilitate superior public health outcomes.

Given that theory monuments our knowledge of determinants of health behavior (Davis, Campbell, Hildon, Hobbs, & Michie, 2015), we should expect to see researchers who incorporate previously established determinants to reference them through theories. It might be the knowledge content (e.g., lists of mediators), rather than the form of knowledge (e.g., theory) that is important for grounding HP research. Further, the extent to which the form of knowledge used in the evaluation of theory usage might drive theory's associations with intervention outcomes is clouded by the variable forms of entities labeled as theory in HP. The use of the term theory in public health refers to a broad and diverse range of conceptual devices (CD). These can include frameworks and models. As discussed in this paper, we should not expect all types of CDs to provide the same benefit. It will be posited that only one type of CD provides all the ascribed benefits of theory. That type is authentic, engrained hypothetico-deductive theory.

The ground can be prepared by recognizing how HP researchers use theory. A common use is as an inventory of constructs that are associated with the outcomes of interest (Glanz, Rimer, & Viswanath, 2015; Lucas & Lloyd, 2005; cf. Bartholomew & Mullen, 2011). Components of these inventories are treated as severable and exchangeable, from and between theories, respectively. Subsetting and mixing of constructs from different theories result in a "collage" of theory components (Goodson, 2010, p. 151). The resultant pot of constructs is often justified by manageability or researcher interest, rather than a clearly articulated deductive consequence of the theory itself (Glanz, Rimer, & Viswanath, 2015). Current recommendations about the use of theory result in the selection of constructs found in one of *many* theories that are available (Achterberg & Miller, 2004; Glanz & Bishop, 2010). Eighty-three options for a theoretical grounding were described in a recent compendium of behavior change theories (Michie, West, Campbell, Brown, & Gainforth, 2014). Deference to the theory that is considered most likely to be correct is not a foremost, or explicit, recommendation for theory selection. Even in a special issue of *Health Education Research* about the use of theory, no researchers justified their implemented theories on evidentiary grounds (Noar & Zimmerman, 2005). Demonstrating the absence of reliable and comprehensive theory of health behavior, the simultaneous use of multiple theories is an explicit recommendation for intervention development (Glanz & Bishop, 2010). Thus, the influence of theory use on an intervention is based on researcher discretion, rather than the deductive structure of *the* theory itself. We suggest that this state is a consequence of how HP has defined theory, in practice.

Objective

A primary goal of scientific inquiry is explanation (Potochnik, Colombo, & Wright, 2018). Theory is the composition of standardized building blocks of science. However, fields such as ecology have been capable of explanation and advancement in areas without a broad matrix of fully-formed strong theories. The aims of this paper are to reconcile the inconsistent conceptual treatment of theory in HP, outline the diagnostic features of one approach to theory classification, and provide an important explanation for theoretical challenges faced by HP. This paper also outlines pathways for scientific advancement in the absence of one critical type of CD.

Because the implications of this paper are broad and complex, an outline of the paper is presented in Figure 2.1. The paper starts by describing the motivations and evidence-base for using theory. A description of the applied definition of theory in HP is then contrasted with the features and diagnostics characteristics of theory in other disciplines. This contrast is developed as a means to respond to the current conceptual ambiguity regarding theory and highlight the difference in conceptualizations of theory between HP and other disciplines. The discrepancies are broad enough to predict and explain deficits in the attainment of the benefits of theory in HP. The paper then explains how ecology is similar to HP and reviews ecology's alternative approaches for conceptual development of the field.

What Motivates Researchers to Use Theory?

There are multiple motivations for incorporating theory in HP. One motivation revolves around the framing of research questions and hypotheses (Krieger & Zierler, 1996). Public health operates at the intersection of social, psychological, and biological processes across multiple ecological scales. Researchers' theoretical lenses are particularly likely to vary at the intersections. It is useful to be aware of, and articulate, underlying assumptions where they may differ from others in the field. Communicating a suite of underlying assumptions that are used to

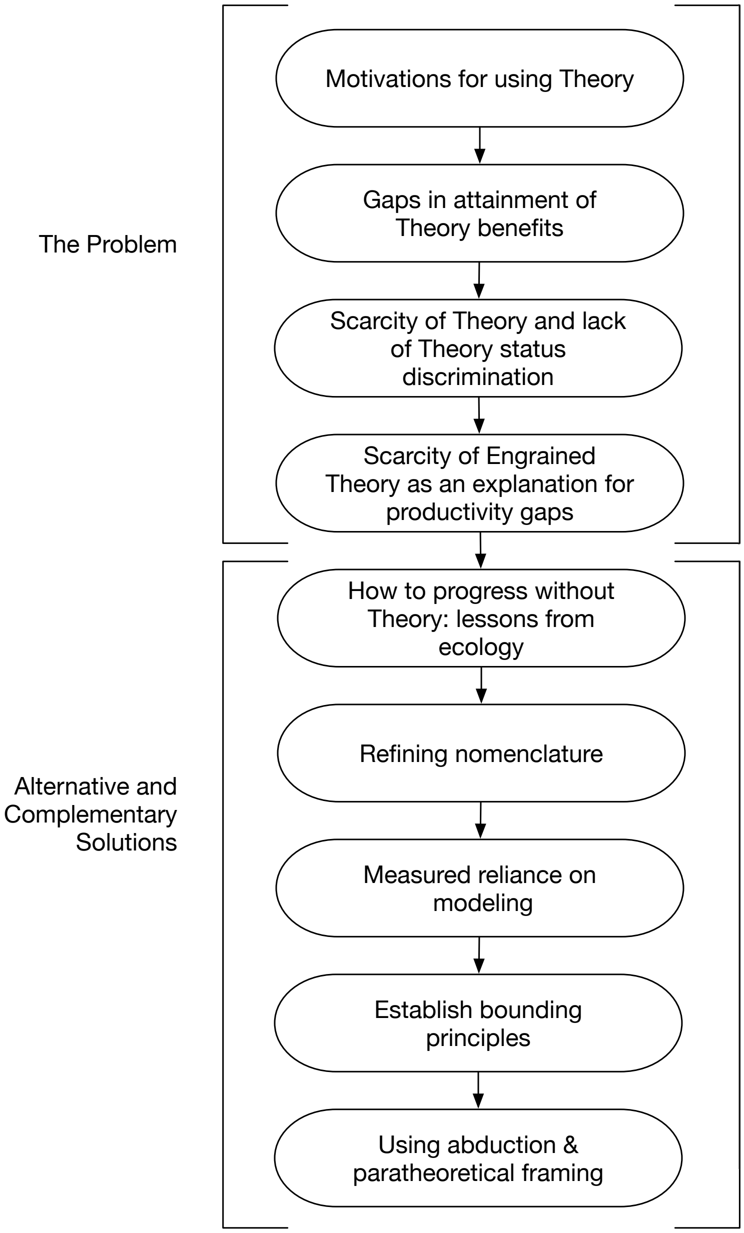


Figure 2.1. Outline of the paper. Rounded boxes are topics of the paper.

justify and interpret a study are made economical and standardized by declaring a cognitive allegiance to a theory (Glanz, 2005).

Perhaps, the most prominent and tangible motivation to use theory is to access a menu of target constructs for measurement and intervention (Crosby, Kegler, & DiClemente, 2009; Glanz, Rimer, & Viswanath, 2015; Michie & Prestwich, 2010; ; Nigg, Allegrante, & Ory, 2002; Noar & Zimmerman, 2005; Michie, West, Campbell, Brown, & Gainforth, 2014). Theories can be distilled to a list of determinants of a health behavior, many of which have a specified position in a causal pathway of a behavioral outcome. These constructs are treated as *hypothesized* mediators of health behaviors (Michie & Prestwich, 2010). It is argued that plugging theory-housed constructs into a study has the benefit of access to a regimented system for intervention design (Michie & Prestwich, 2010). This can be facilitated through the use of matrices of intervention components that map onto specific theory-based constructs (Michie et al., 2008). Thus, theory facilitates understanding of related behaviors, providing targets and strategies for interventionists (Bartholomew Eldredge et al., 2016; Glanz, 2005; Michie et al., 2008).

A related benefit of using theoretical constructs is a more productive field. It is argued that the inclusion of these theory-housed constructs facilitates the evaluation of theory, which improves the empirical basis for theory generalization and for making theory amendments (Michie & Prestwich, 2010). The resultant refinement of theory further strengthens the field's knowledge base (Glanz, Rimer, & Viswanath, 2015). Changes in theory can lead to changes in future research directions and clinical practice (Glanz & Bishop, 2010). However, it is unclear why the *form* of knowledge currently identified as theory in HP is critical for the field to advance. Theory can provide a reference point for comparing findings and a way to standardize the field’s knowledge base, but this is also true for other types of CDs (e.g. conceptual framework)(Crosby, Kegler, & DiClemente, 2009).

Beyond housing a pool of related constructs, Goodson provides a set of benefits for “theoretical thinking” (Goodson, 2010, p. 23-30). Pivoting one’s approach on theory provides: perspective, guidance on ethics, defense against hegemony, provision of direction for, and structuring of, research activities. It also builds the scientific knowledge base. Obtaining blueprints for prediction and control of health-related outcomes is also included in Goodson’s general presentation (Goodson, 2010).

A central aim of science is the construction of theories (Achinstein, 1968). In many scientific fields, attainment of trusted scientific theory has resulted in the availability of engineering tools useful for interventions. Such tools would be eminently useful in HP. Given the limited resources and time sensitivity related to many public health concerns, running screening studies on multitudes of intervention strategies for every problem is not feasible. The aspiration is to hold a theory that informs us as to which tools to apply, without having to start from scratch for every problem that arises. Harre wrote, “Theories are the crown of science, for in them our understanding of the world is expressed" (Harré, 2004, p. 75). Unfortunately, this statement might not apply to HP. As will be discussed below, the *form* of HP theories differs from other sciences.

What is the Evidence Supporting the Use of Theory?

Based on a body of literature reviews published in the first decade of the 2000s, it was concluded that the application of theory was beneficial (Glanz & Bishop, 2010). However, a newer synthesis is more equivocal (Michie, West, Campbell, Brown, & Gainforth, 2014). A recent meta-analysis of physical activity interventions also failed to find support for theory-informed approaches (Lock, Post, Dollman, & Parfitt, 2020). This has sparked a debate about the utility value of theory in health promotion (Hagger & Weed, 2019). Current findings primarily rely on assessing patterns *across* comparative studies; HP's core body of primary research is not designed to evaluate the use of theory (cf. Reback, Fletcher, Shoptaw, & Mansergh, 2015). The reliance on indirect evidence, combined with the indeterminate usage of theory, results in several deficiencies in the evidence base available for evaluating theory's benefit. The use of theory may be associated with the quality of the study design or intervention design features not addressed by theory. Among reviewed studies of cancer screening, all studies failing to invoke theory in the research report scored the lowest on a structured study quality scale (Noar et al., 2007; cf. Albada et al., 2009). There is also potential confounding between projects that invoke theory and the type of care that is taken to construct and execute an intervention (Glanz & Bishop, 2010). Assessment of benefits for modifying ultimate behavioral targets is further clouded because no accounting is made for differences in the potency of interventions to modify directly targeted constructs. For some health behaviors, the success of the intervention is related to dosage (e.g., Haller et al., 2016). Further, the same content delivered via different modalities or interventionists can affect potency (Cadigan et al., 2015; Carey, Scott-Sheldon, Carey, & DeMartini, 2007; Meier, Barrowclough, & Donmall, 2005; Project MATCH Research Group, 1998). Inclusion of targeted and intermediate mediators in the measurement regime of an intervention study can assist in gauging the potency and play a role in explaining the success, or failure, of an intervention (Nigg et al., 2002; Rothman, 2011). The use of theory does not immunize an intervention study from poor design or execution, so comparisons across studies presents interpretation challenges.

Conceptual imprecision creates an additional set of problems. It has been typical to define the use of theory based on whether or not a theory was invoked in the article (Albada et al., 2009; Ammerman, Lindquist, Lohr, & Hersey, 2002; Gardner, Wardle, Poston, & Croker, 2011; Noar et al., 2007; Kim, 1997). Research reports rarely include an adequate description of the nexus between theory and specific features of the intervention. Notably, structured descriptions are becoming more common (Bartholomew Eldredge et al., 2016; Peskin et al., 2017; Pot et al., 2018). The most common link between design and theory is via targeting of constructs found in a named theory. However, utilized constructs typically represent only a subset of a theory and are sourced from multiple theories (Albada et al., 2009; Noar et al., 2007; Michie, Jochelson, Markham, & Bridle, 2009; Prestwich et al., 2014). Notably, some studies without reference to any particular theory utilize constructs that are found in theories. In a review of print communication interventions, 96% of studies included concepts that authors traced to a referenced theory, although not always to the correct theory (Noar et al., 2007). Of the complement group of studies, 60% used a concept that qualified as a theoretical concept even though an associated theory was not described. In Noar’s review (2007), the mean number of concepts was four, which indicates that most studies use fewer constructs than in the referenced theories. In another review, only 9% of studies were found to have used all the constructs of a theory in an intervention (Prestwich et al., 2014). There is little indication that selected constructs are solely deduced as the appropriate targets of intervention through theory application. Inadequate descriptions of how theory was used to inform the choices made in the engineering of an interventions has been previously noted (Albada et al., 2009; Gardner et al., 2011; Hagger & Weed, 2019; Michie et al., 2009; Michie & Prestwich, 2010). There may be no difference in intervention techniques among some projects invoking theory and those that do not

(Michie et al., 2009). There are reports of researchers admitting to only create a veneer of theory incorporation, just to please reviewers (Goodson, 2010). If there was no substantive integration of theory in the research used to gauge theory's benefit, conclusions about theory's benefit should be considered tenuous.

Even among reviews reporting at least one supportive association between theory usage and outcomes, effect sizes have been small, inconsistent, only marginally significant, or questionable upon inspection of the interval estimates (Ammerman et al., 2002; Kim, 1997; Noar et al., 2007; Taylor, Conner, & Lawton, 2012). This obliges us to be cautious in predicting that HP theory, in its current form, will substantially raise the tide of intervention effectiveness. The lack of large effect sizes could be a result of the attenuation caused by poor measurement of theory use and weakness in how experimental comparisons are constructed. More importantly, the field's propensity to wrap its knowledge into theories means that there may be few opportunities to cleanly compare an intervention that is informed by prior research without its evidence-base coinciding with knowledge that is monumented in published theories; for example, newly identified risk-factors may be added to a theory. Another consideration is the paucity of evaluation and discussion of the specific theory used in studies. A functional definition of theory that allows any CD to suffice means suboptimal CDs will dilute the effects of stronger CDs, such as a severely tested theory, when their effects are pooled (see Chalmers, 2013; Chalmers, 2010; Mayo, 1991; Popper, 2014) for discussions about types of severe tests). This is underappreciated. There is recognition that some theories may be inappropriate for the contexts in which they were applied, but the verisimilitude of the candidate theories is not adequately considered (Gardner et al., 2011; Michie et al., 2009). Some of the cloudiness regarding the evidence for the benefit of theory use in interventions can be addressed through a more deliberate structuring of the way CDs are categorized, used, and analyzed.

CONCEPTUAL FOUNDATIONS

Much of the theory-related vocabulary is ambiguous in HP (Bartholomew & Mullen, 2011). There are several HP related expositions about theory that provide definitions of scientific theory (Glanz & Bishop, 2010; Glanz, Rimer, & Viswanath, 2015; Goodson, 2010; Nilsen, 2015). These varied treatments can be partially reconciled, but they poorly align with the divergent use of the terminology. For example, theory is sometimes used as an umbrella term for all CDs, a term for a specific CD type, and a term used for causal explanations (Datta & Petticrew, 2013; Hawe, 2015). Models are inconsistently treated as distinct from theory. Frequently, the discussion of theory does not clearly distinguish which recommendations apply equally to CDs titled model and those titled theory, in HP (Glanz & Bishop, 2010). Fidelity to a discriminatory taxonomy of CDs will aid evaluation and provide clarity to recommendations related to different types of CDs. For the remainder of this paper, CD classes will be differentiated. When referring to a specific type of CD known as Theory, it will be denoted with a capital T (Gray, 2017) to distinguish it from its looser use that includes all abstracted CDs–including models.

A prevalent way the term theory is used in HP is described by Glanz et al. (Glanz, Rimer, & Viswanath, 2015). Citing Bandura, theory’s function is described as cataloging determinants of health behaviors (Bandura, 1986; Glanz, Rimer, & Viswanath, 2015). It is recognized that this construction fundamentally differs from other sciences (Glanz, Rimer, & Viswanath, 2015). Here is the problem; if the construction differs, the benefits will also differ. So, what are the differences?

In this paper, Theory is a hypothetical representation designed as a conceptual tool for structuring our thinking about phenomena (Boniolo, 2007). It is a device with specific characteristics that is used to represent scientific knowledge. The American Association for the Advancement of Science defines a scientific Theory as a well-substantiated explanation of some aspect of the natural world, based on a body of facts that have been repeatedly confirmed through observation and experiment (American Association for the Advancement of Science, 2005). Theories are neither conjecture, speculation, nor weakly supported. Thus, we should treat Theories as potentially accurate maps of the world (under a realist perspective). They should impart understanding that is believed likely to be true (Achinstein, 1968; Harré, 2004). The portable knowledge of Theory can be leveraged as the basis for novel engineering approaches (Hacking, 1983). Theory of this sort could successfully direct HP intervention design.

At a different level, Theories are also described as abstract representations expressed as formal deductive systems that predict natural events (Thompson, 1989). They are large-scale systems of ideas about natural phenomena. They are also more general and elaborate units of knowledge than a typical hypothesis encapsulates (Potochnik, Colombo, & Wright, 2018). Theories provide general and formal means of organizing known empirical patterns, allow for prediction of future empirical observations, and give meaning to those patterns (Boniolo, 2007). Theory is also considered a source of solutions to problems, in the form of explanations for the various patterns associated with a specified type of phenomenon (Harré, 2004; Popper, 1994).

Theory outlines the constituents, their interactions, and mechanisms of the processes that result in the patterns. This is more than a collection of constructs connected by a limited set of qualitative associations. Theory provides explanations for observable phenomena, provides structure that grants predictions about those phenomena, and delineates the scope of applicability. General features of scientific Theories include: 1) a set of interrelated propositions, concepts, and definitions, 2) a specific domain of applicability, 3) a description of relationships between components, 4) mechanistic explanations for the relationships, and, consequently, 5) predictions about the behavior of the system. Although this list is more inclusive than reported by others in the social sciences, it is consistent with their intent (Imenda, 2014; Wacker, 1998). Definitions developed specifically for health behaviors research focus on feature 3 and 5 (Davis et al., 2015), which are also consistent with other forms of scientific representation. It should be recognized that not all CDs identified as Theory clearly contain all five attributes or other diagnostic criteria described below.

The features of Theory are not settled. There has been substantial philosophical dialogue about what should constitute a Theory (Winther, 2018; Lambert & Brittan, 1992). This may, in part, reflect the flexibility that scientists provide in the identification of Theory. However, the benefits attributed to Theory across the sciences are based on well-formed Theories, which meet the criteria presented here. This point should be considered when adopting working definitions of Theory in HP.

Falsifiability is the most distinguishing diagnostic feature of scientific Theory. Theory should, in principle, provide clear opportunity to identify distinct predictions about future empirical observations that would trigger the demotion of the Theory if borne out (McMullin, 1976). Naïve methodological falsification is achieved if investigation of the types of features and patterns used to generate and justify the Theory later present patterns leading to the Theory's refutation. Sophisticated methodological falsification relies on the capacity of the Theory to deductively predict *novel* facts about nature. Novel patterns are the knowledge created by

Theory–Theories are more than devices for capturing accumulated knowledge. Deductively discovered knowledge is a chief benefit of Theory.

Theories are not amenable to universal verification, so falsification tests are our strongest tools in solidifying our commitment to a Theory. Falsifiability is so pivotal that it has been considered a demarcation between science and pseudoscience (Mayo, 2019; Popper, 1994; Staley, 2014). Taking an alternative view that science operates instead on gauging the verisimilitude of Theory can be compatible with accepting that some explanations can be wrong, and in need of discarding. Theory is more than a prediction device. It is explanation that paves a way for understanding how our world operates. Public health infrequently emphasizes Theory's falsifiability (or verisimilitude) as a critical feature (Noar & Zimmerman, 2005; Michie, West, Campbell, Brown, & Gainforth, 2014; cf. Goodson, 2010). Rarely are specific novel predictions of HP CDs identified, and even more rarely are they severely tested. As described below, this may be a consequence of an ambiguous conception of theory in HP and the sparsity of CDs applicable to HP that produce novel deductive predictions.

Theory Form: How Do Philosophers View Theory?

Although scientists need not adopt philosophers’ perspectives (Murray, 2001), there are benefits in reviewing their contributions. Different philosophical perspectives on the idealized form of theory include the: syntactic, semantic, and pragmatic views of Theory (see Appendix). It is unclear if most CDs used in HP conform to these views. It is clear that Theory, under the syntactic or semantic view, is not a collection of antecedents of a health behavior, which is sufficient to function as theory in HP.

Philosophical analysis is grounded in the details of how science is practiced; however, philosophers focus on features related to the structure of Theory in the natural sciences. On the other hand, scientists are more interested in building knowledge than delving into the abstract structural analysis of scientific representation. Nonetheless, it can be useful for us to consider structure. For example, the philosophical views remind us that Theory should self-describe where and why it is applicable. For instance, formalized inclusion of the demarcated limits of the Theory's domain of application is a key feature of Theory identified by philosophers. Explicit description of the class of phenomena that a CD is applicable is often overlooked, or weakly developed, in CDs branded as Theory in HP. Under the semantic perspective, the absence of crisp information to this point would disqualify a CD as Theory.

Features of Theory: How Do Scientists View Theory?

What should we look for in HP Theories? The presentation of the diagnostic features of Theory includes only some of the characteristics associated with well-formed Theory. A comprehensive suite of characteristics is provided here by combining typical exemplar descriptions (Table 2.1). The most common descriptions of Theory indicate that it is explanatory and predictive. Explanation requires knowledge of the causal mechanism. Theory is discrete in that it is built out of context-independent elements, where features of Theory structure (not content) are devoid of cultural influence, including traditions and values. Ideally, well-formed Theory can be decoded in the same way by all persons, regardless of their cultural lens. Theory is systematic, such that it describes consistent relationships among *all* of the components. It is complete where Theory specifies the effect across the components' ranges (Meehl, 1990). Theory is predictive in that its consequences can be verified by empirical observation (Flyvbjerg, 2001). There is some recognition that the predictive properties should be unificatory, such that well-formed Theory can predict a variety of phenomena. Natural scientists also describe Theory as being adequately described to eliminate ambiguity. Its propositions should be clear.

Falsifiability is often omitted from feature lists. It is often treated implicitly as the overarching diagnostic feature of science, and its Theories. We treat falsifiability as a hallmark of Theory.

Table 2.2. Desiderata of archetypical hypothetico-deductive Theory

	Suites				Example Cognitive Devices	
Author	Colyvan	Achinstein	Flyvbjerg	*Authors' Synthesis*	Theory of Planned Behavior	Optimal Foraging Theory
Field	Ecology	Philosophy	Social Science	Interdisciplinary	Public Health	Ecology
Characteristic	Explanatory	Imparts Understanding	Explanatory	Explanatory	Moderate	Moderate
		Propositions explicit	Explicit	Propositions explicit	Moderate	High
	Unificatory¥		Predictive	Unificatory¥	Low	Moderate
			Complete	Complete	Low	Moderate
		Applicable context*		Delineated	Low	Moderate
	Fruitful				High	High
		Nonderivative		Nonderivative	High	High
		Elemental		Elemental	Low	High
	Falsifiability		Falsifiable*	Refutable	Low	Moderate†
		Believed might be true; Not known to be false		Potential to be considered approximately true	Low	Moderate
			Universal	Universal	High	High
			Discrete	Discrete	High	High
			Systematic	Systematic	Low	High
	Elegance				Moderate	Moderate

Note. Not all treatments precisely aligned. Characteristics are listed on the same row where there was substantial overlap. Some terms were modified from an author's original presentation to aid comparison. Exemplars are evaluated on their strength of the synthesized list of characteristics. *Some characteristics were not explicitly included in authors' listings but were included in their description of science methodology. ¥ Accommodation of prior known empirical patterns is consolidated with forecasted patterns. †A broad protective-belt (Lakatos, 1968) has protected the core of optimal foraging theory (Pierce & Ollason, 1987; Pyke, 1984; Perry & Pianka, 1997; Scheiner & Willig, 2011b; Pyke, 2019; Pyke, 1984). Although it has been most consistent with qualitative findings, useful empirical knowledge about foraging behavior has accumulated through iterations of interleaving model building and experimentation (Pyke, 2019; Werner & Mittelbach, 1981).

Several features that are particularly relevant to the consideration of leading HP CDs were highlighted by Achinstein (1968; Table 2.1). The first is that Theory is not directly derivable from another theory without at least non-trivial integration of additional premises; Theory is nonderivative. Theory is elemental; it is distinct from other Theories, nonseverable, and not an amalgamation of other Theories' components. Theory is something that can be believed to be true and not known to be false. Being a useful instrument for control and prediction is not enough. Instead, we treat it as if it possesses truth value. Consequently, Theories are neither modular nor interchangeable alternatives.

There are several features often left out of descriptions of Theory, but may have been omitted because they are implicit. For example, Theory should be universal. The Theory should apply in all places, times, and contexts. As mentioned above, another common omission is that Theory has a specific domain of applicability (Staley, 2014). This is particularly relevant in the study of complex systems. When does a Theory apply? A well-formed Theory will ably self-describe the delineation of its domain. A common challenge in HP is determining which CD should be applied.

Finally, Theory should include a set of propositions, where the propositions include central, distinctive, and integrated assumptions. The central propositions are the minimally sufficient set that are required for the Theory to function. Theory is a set of *consequences* of some specified axioms. The integrated assumptions may be consequences of other Theories, but a Theory's propositions distinguish it from all other Theories.

Why walk through all these features of Theory? To highlight the point that Theory is much richer than an inventory of antecedents, which is currently accepted as an operational definition of theory in HP (Glanz, 2005; Goodson, 2010, p. 151). The next section will outline how another scientific discipline dealt with similar incongruities.

FEATURES OF THEORY IN ECOLOGY AND HEALTH PROMOTION: A HISTORICAL PERSPECTIVE

A parallel between HP and ecology exists in the absence of a consolidated and consistent definition of theory (Gorelick, 2011; Krebs, 2017; Marquet et al., 2014). In both fields, the polysemy blurred the understanding of Theory. In ecology, even champions of theoretical ecology have been unsure of the field's definition (Harper, 1980; McIntosh, 1980). In some contexts, ecologists describe Theory as a scheme or system of ideas that accounts for a group of phenomena (Harper, 1980). This is consistent with the description presented above but is open to more varied and broader construction. For example, terminological usage has included *hypothesized* explanations (Smith, 1976). In ecology, theory can also refer to all academic activities involving mathematical or computational modeling (Kolasa, 2011). The specific intent of the language can often be identified by its context (Hodges, 2008). Nonetheless, the lexical ambiguity may have muddied expectations about CDs in ecology, as it has in HP (see Scheiner, 2012).

Both Ecology and HP have demonstrated an inconsistent terminological and conceptual blurring of the distinction between models and Theory. This impacts the attitudes of practitioners toward theory. Wildlife managers did not see ecological theory as providing reliable tools and predictions (Romesburg, 1981). Indeed, hunches and common sense were seen as equal to theory in the design of interventions (Romesburg, 1981). However, this view was based on the terminological usage of theory as being synonymous with unparameterized models that are unconnected to well-formed Theory. The view of theory is revealed by statements such as, “Modeling was never intended to function as a means to scientific knowledge" (Romesburg, 1981, p. 310). We recognize that the function of modeling grossly contrasts with Theory's capacity to monument and create knowledge. Models and Theory are very different (Boniolo, 2007). As with some HP researchers (Nilsen, 2015), some ecologists have recognized this distinction. Ecologists have also understood that not all useful explanations constitute a Theory.

Variance in the treatment of theory is also recognized within the HP literature (Glanz, Rimer, & Viswanath, 2015; Goodson, 2010). Where Theory as a specific CD is defined, its features include description, explanation, and prediction (Glanz, Rimer, & Viswanath, 2015; Goodson, 2010). However, “sets of tested empirical generalizations”, “theoretical orientations or perspectives”, and “various types of ideas, speculations, hypotheses, models, criticisms, conceptual frameworks, or any propositions interconnected with words (and even scholars’ personal beliefs) are sometimes called theory” (Goodson, 2010, p.7). HP doesn't consider its theories to be systematic; use of a particular theory can cause a "blinding effect," where a researcher fails to consider important processes that are absent from the chosen CD (Goodson, 2010, p. 86). Well-formed scientific Theories are less vulnerable to this effect because they are systematic and complete; they contain what they need.

In terms of benefits, CDs titled as models are often treated as equivalent to those titled Theory in HP. One response to this state is Goodson’s view of theory, where theory is pluralistic, instrumental, and intimately connected to the process of explanation. Here, CDs resulting from the *process* of answering questions of causal explanation are theories. Goodson weaves a pragmatic approach by considering context and meaning with the development of explanations

(Goodson, 2010). The power of theory is then tied to the power of explanation. However, the resultant CDs may not possess the features of Theory presented here (Table 2.1). Again, not all forms of explanations provide the same benefits. If the intent is to consider any explanation as a Theory, the proposed benefits should match this looser construction.

An exemplar from HP is now considered. The Theory of Planned Behavior (TPB) has features that map to Theory, but many are deficient (Table 2.1). This does not discount TPB's contribution to HP (Montano & Kasprzyk, 2015). The intent of TPB was to provide an explanation for most behaviors of interest to social psychologists (Ajzen, 2011). The broad edges of this putative domain of application are not entirely explicit, nor does the TPB self-describe when it is applicable (without creating a tautology). This is common among behavioral CDs and results in broad application beyond CDs' original development contexts (Hagger & Weed, 2019).

TPB's structure has changed in recognition of limitations to behavioral control. This why the perception of control construct was appended to the Theory of Reasoned Action to arrive at the TPB (Ajzen, 2011). Ajzen (2011) also recognized that actual control should be considered part of the causal system and that the perceptions construct *could* be used as a proxy. The fuzziness regarding which (or when a) control construct should operate reduces the explicitness of the theory. This is also an admission that the CD is not systematic, because it is implied that the relationship between actual control and perceptions varies in ways that are not described by the theory. The simplest indicator that TPB is not a well-formed theory is that it is not elemental, because the Theory of Reasoned Action is a subdivision of TPB.

TPB holds behavioral intentions as the proximal driver of behaviors (Ajzen, 2011). Implicit in the TPB is that cognition drives an intention mediated process leading to goal-oriented behavior (Ajzen, 2011). This leaves out what we know about the influence of unconscious processes on many behaviors (Kahneman, 2011; Chaiken & Trope, 1999). Although dual-systems influencing behavior may be mediated by the same constructs (Ajzen, 2011), this is not a feature of TPB. Substance use researchers are particularly attuned to the role of non-cognitive behavioral influences (Koob & Volkow, 2016; Norman & Conner, 2006; Norman, 2011; Pulido, Brown, Cummins, Paulus, & Tapert, 2010), which can profoundly conflict with distal behavioral intentions. Developmental researchers are also forced to recognize the keystone role of self-regulation (Miller, Lo, Bauer, & Fredericks, 2020), which can be requisite for planned behavior to manifest into action. Further, affective processes are known to influence behavior (Kiviniemi et al., 2018; Lerner, Li, Valdesolo, & Kassam, 2015; Lerner & Shonk, 2010; Mendl & Paul, 2020). Most health behaviors of concern are likely to be influenced in part by processes not addressed in the TPB, even those that are dominated by deliberate planning. As with many health behavior CDs, the structure of the TPB appears to only capture one branch of the complex network of behavioral influences in which it is embedded. Well-formed theories address the whole of interrelated phenomena.

Critically, Theory should describe and explain to which phenomena it is applicable, and that’s not adequately developed for consistent application in HP. As a Theory, TPB is certainly not unificatory, as it is a CD that is limited to those behaviors for which it ends up describing, without linking otherwise seemingly unrelated phenomena. Even where behavior appears to be overwhelmingly driven by TPB’s cognitions, behavioral intention, and control constructs, the TPB is not considered by the field as likely to be true, systematic, or certain to outcompete all other CDs for a clearly defined class of behaviors. Thus, in HP the use of TPB is treated as discretionary. If it were considered as likely correct or superior, its use would be expected, rather than an optional alternative.

TPB includes a list of external factors, such as demographics, which do not have clearly defined relationship structures with other features of the Theory; this another indication that TPB is not systematic. Worse, these external factors have been found to operate outside the pathways described the TPB (Sniehotta, Presseau, & Araújo-Soares, 2014). Further, none of the relationships are quantified for any part of their range (Meehl, 1990). As with most health behavior theory, it is unclear to what degree TPB provides *unique* quantitative, or even qualitative, prediction or control. There are some causal paths in this model, yet an interventionist would not know which is the most crucial, effective, or efficient to intervene upon, because the strength of their effect is not specified (Rothman, 2004). There are no risky novel predictions that emerge from TPB, which has resulted in it being considered unfalsifiable (Greve, 2001; Ogden, 2003; Smedslund, 2000). This reflects its semi-amorphous nature and application. Its falsifiability has been claimed under specific formulations (Trafimow, 2015). However, under this view, it has already been falsified (Trafimow, 2015). TPB does not provide a systematic, delineated, and fully explicit account of a range of specified phenomena.

Despite calls for its abandonment (Sniehotta et al., 2014), TPB can provide service to HP, but it will not provide the benefits of a well-formed Theory. Deficiencies of CDs treated as Theory may help explain the generalized discounting of benefits for *all* theory in HP. The most common deficiency is an inadequate description of the CD's domain of application. TPB does not work for many health behaviors to which it has been applied (Sniehotta et al., 2014). Is that because TPB is not true or because TPB was not intended to be applied to most health behaviors? Practitioners feel unguided as to when a particular CD is to be used, but a well-formed Theory should answer that question.

HP is not alone in failing to produce CDs with articulated domains of application. Lack of domain delineation is also a common deficiency in ecology (Kolasa, 2011). In both ecology and HP, the term theory is used to refer to the investigation and communication of abstract knowledge. Both fields consider model building as a part of the disciplinary arm conducting work related to theory. Considering how the two fields differ in traditions of modeling provides important insight. Ecologists have built a vast literature on the behavior of mathematical and computational models. There is less analytical (or even comparative) investigation of models in HP (cf. Cepeda et al., 2018; Homer & Hirsch, 2006; Levy, Bauer, & Lee, 2006; Sterman, 2006; Strathdee et al., 2010). Ecologists have also been more likely to recognize that their field is only partially covered by strong well-formed descriptive Theory, even where numerous alternative models abound. This is because ecology makes a stronger distinction between models and Theories than HP. Ecologists also distinguish a set of benefits specific to modeling, which are present even in the absence of Theory (Renshaw, 1993). HP can also make these distinctions and leverage the value of modeling, even where Theory is scarce.

Different types of CDs should not be expected to achieve the same success. Those strong in the features of Theories presented here should be expected to provide stronger benefits related to the development of interventions. Both ecology and HP have experienced skepticism among its practitioners regarding the benefits of Theory to support their efforts; however, practitioners in both fields may be confusing Theory with other types of CDs. The same can be said about making the distinction between an Inchoate Theory and one which has strong verisimilitude. In HP, this may be a result of the scarcity of well-formed strong Theory that can be used to create contrasts against other CDs.

It has been suggested that Theory, as construed in the natural sciences, may not be attainable in the social sciences (Flyvbjerg, 2001). Consequentially, it might be counterproductive even to use the word theory in social sciences (Flyvbjerg, 2001). Ecology's experience indicates that HP may need an alternative to this suggestion; HP may be unlikely to eliminate, or even reform, terminology in current use (Hodges, 2008). Instead, the field can benefit from extending its vocabulary to communicate univocally and clearly distinguish among types of CDs and their evidentiary status. To avoid confusion in this paper a tentative taxonomy of Theory is presented in Table 2.2. The consequences of adopting a refined taxonomy will improve our recognition that Engrained Theories (see Table 2.2) are sparse in HP. For example, TPB is one of the most widely used health behavior CDs (Davis et al., 2015), but is not considered approximately true and holds no domain where it is exclusively recommended. So, in addition to this leading CD not holding the features of a well-formed theory, its evidentiary status (as a Theory) is tarnished (Sniehotta et al., 2014; Trafimow, 2015). Unfortunately, there are no Engrained Theories that are part of HP's current suite of CDs. As will be elaborated upon below, ecology demonstrates that progress can still be made even if Theory is sparse.

Theory Status: How Should Theories be Differentiated?

The identification of the developmental status of Theory has been recommended in ecology (Kolasa, 2011; Pickett, Kolasa, & Jones, 2010). Applying a nomenclature communicating the status requires the recognition that Theories vary in their merit. This is important because we should have expectations commensurate with their status, and our research programs can be optimized to focus on questions that are most appropriate for the stage of a Theory's development and evaluation (Pickett, Kolasa, & Jones, 2010).

Table 2.2. Provisional taxonomy of scientific theory.

Conceptual Device	Description
Archetypal Hypothetico-Deductive Theory (H-D Theory):	Predictive and explanatory scientific Theory with properties presented in Table 2.1, which are consistent with those attributed to Theories in the physical sciences. Physics provides examples of strong H-D Theory with salient novel predictions. Examples include the Higg's boson and the recent black hole measurements (Aad et al., 2012; Event Horizon Telescope Collaboration et al., 2019).
Emerging Theory	H-D Theory that has gathered empirical support as a potential replacement for prior Theory or addresses a domain not previously covered by Theory, however, the body of evidence is not yet conclusive.
Inchoate Theory	H-D Theory that has been merely proposed or evolved with little work aimed at elucidating and challenging its empirical consequences. It should only be considered an incipient or prospective explanation. Scientific skepticism would dictate that it not be adopted for application out of hand, even if it addresses a gap in the theoretical landscape.
Engrained Theory	Substantiated H-D Theory so firmly accepted by the field it is considered a working fact (Lewis, 1982). These should have surmounted strong empirical challenges and hold high verisimilitude to the extent that it accepted as the exclusively correct explanation (McMullin, 1976). Anomalies (Staley, 2014) and viable alternative Theories should be absent.
Tarnished Theory	An otherwise Engrained Theory that has not been discarded but is at risk of degenerating because of new anomalies or challenges posed by a new competing Theory. Outright rejection may not be warranted, but its prior verisimilitude is in question.
Constituent Theory	H-D Theory which is constrained by a set of principles that apply more broadly than the Theory itself (Scheiner & Willig, 2008).
Organizing Framework	Conceptual device that might not be predictive, explanatory, or refutable as a whole. It is any device used to structure investigations, interpretations, and the communication of assumptions and findings. It may include some explanation regarding the patterns observed, but is materially deficient in diagnostic aspects of H-D Theory (Flyvbjerg, 2001). This CD includes conceptual (Imenda, 2014), theoretical (Pickett, Kolasa, & Jones, 2010), or determinant (Nilsen, 2015) frameworks.

Where there have been calls for interventionists to increase the use of theory, there have been responses that theory is not seen as useful (Rothman, 2004). Practitioners' concerns about theory can be partially attributed to holding expectations that any CD, of any status, will suffice. It may be that the dissonance can be alleviated by clarifying if recommendations are limited to Engrained Theory. Strong Engrained Theory needs no salesperson. HP will benefit from using the posture of the physical sciences where Theory is treated as the CD that can hold the greatest truth value. Whereas ecology has also recognized the importance of reserving credence for Engrained Theories with high verisimilitude (Smith, 1976), HP provides a contradictory and potentially confusing view of theory. Interventionists have even cited theory as the lowest tier of evidentiary support (Campbell et al., 2000). Adopting an expanded taxonomy can help reconcile these cross-disciplinary discrepancies and add clarity to recommendations made to practitioners. Rather than recommending the use of *any* Theory in HP practice, the recommendation would quickly transform into promoting the use of Engrained Theory.

HISTORY OF THEORY IN ECOLOGY: WHAT CAN WE LEARN FROM ECOLOGY?

The history of ecology holds several lessons and opportunities for HP. The tension within ecology regarding its theoretical development was outlined by McIntosh (McIntosh, 1986). There are four relevant historical themes: 1) the substantial industry of mathematical modeling of ecological processes, 2) field ecologists' skepticism of theoretical work, 3) recognized scarcity of Engrained Theory, and 4) ecologists' circumspection of their ability, or need, to achieve idealized physics-like Theoretical matrices to buoy and direct the field. An equal counterpart to the first theme is generally absent in HP (cf. Homer & Hirsch, 2006; Levy et al., 2006; Smit et al., 2011). The second theme is similar to some practitioners' attitudes in HP. The last two contrast with the dominant attitudes in HP. These themes and their relationship to HP will be discussed below.

The Mathematical Ecology Industry: Modeling without Theory

Mathematical modeling has been employed extensively in ecology. This effort is the core of the work, often referred to as theoretical ecology, particularly in the context of population dynamics. Some of this work has illuminated the potential for complex dynamics arising from systems with simple structures, such as idealized interactions between two species. Lotka-Volterra models demonstrated how competing species could exclude each other from a location, where they would otherwise occur in the absence of their competitor (Wangersky, 1978). Simple models for predator-prey interactions also demonstrated situations in which interacting populations could result in stable dynamics or extinction (May & Oster, 1976; May, 1976). The modeling required ecologists to be explicit, and quantitative, in their assumptions. Consequently, it provided checks on cognitive errors regarding the results of ecological interactions, provided machinery for making quantitative predictions, and focused ecologists' thinking about substantive problems (Christiansen & Fenchel, 2012). Even though modeling helped ecologists think and orient, the foundational deterministic modeling was seen as a dubious approach to quantitative prediction (Oster, 1981; Smith, 1952).

Where the early modeling was pressed to predict real populations, such as small mammal population cycles, there were a series of failures (Sagoff, 2016; Chitty, 1996). This was, in part, attributed to the models reflecting an incorrect explanation for the observed dynamics (Boonstra, Krebs, & Stenseth, 1998; Chitty, 1996; Sagoff, 2016). The models were not, nor a constituent of, Engrained Theory. On the other hand, there have been modeling successes. Modeling developed closely with observational data has been useful in managing ecological resources, like fisheries and threatened species (Akçakaya, Franklin, Syphard, & Stephenson, 2005; Beamish & Rothschild, 2009; Converse & Moore, 2013; Hilborn & Ovando, 2014; Lawson, Regan, Zedler, & Franklin, 2012). Models used to inform ecological interventions are commonly phenomenological models (Figure 2.2). Ecologists know these models are not derived from the deductive consequences of Theories, and expect commensurately less of them (Schnute & Richards, 2001). Natural resource managers assume the models provide better service to decision making than deliberating without the quantitative analysis of their assumptions. However, the phenomenological models are recognized as imperfect and their application benefits from empirical monitoring, model adaptation, and substantial caution. Patterns generated by models can also help illuminate possible mechanisms behind phenomena (Collie, Richardson, & Steele, 2004).

Theory does not monopolize hypothesis generation. Important lines of inquiry have been conducted as part of exploratory research where important gaps in ecology's knowledge base existed (e.g. Oechel et al., 1993). Models that are simplifications can form conceptual seeds of motivating hypotheses about how ecosystems work. In ecology, a suite of basic models often provides a skeleton on which supplemental features, such as time lags, stochasticity, and spatial and temporal heterogeneity, are added (Kolasa, 2011). This further helps ecologists guide the development of their understanding. These models are undeniably a reflection of how ecologists think. They also provide explicit and precise scientific representations that aid in communication, evaluation, and comparison. The modeling catalyzed research, but empirical knowledge of the specific system of interest has also been a generator of questions and solutions (Kolasa, 2011; Shrader-Frechette, 2008). Ecology demonstrates that models can provide some services sought from HP theory.

The perceived danger has been that the models themselves become reified (Krebs, 2017). Ecology has wrestled with models disconnected from strong empirical evaluation being mistaken for representations of reality adequate for use in prediction, control, and understanding (Romesburg, 1981). To some extent, plainly oversimplified models were expected to operate like

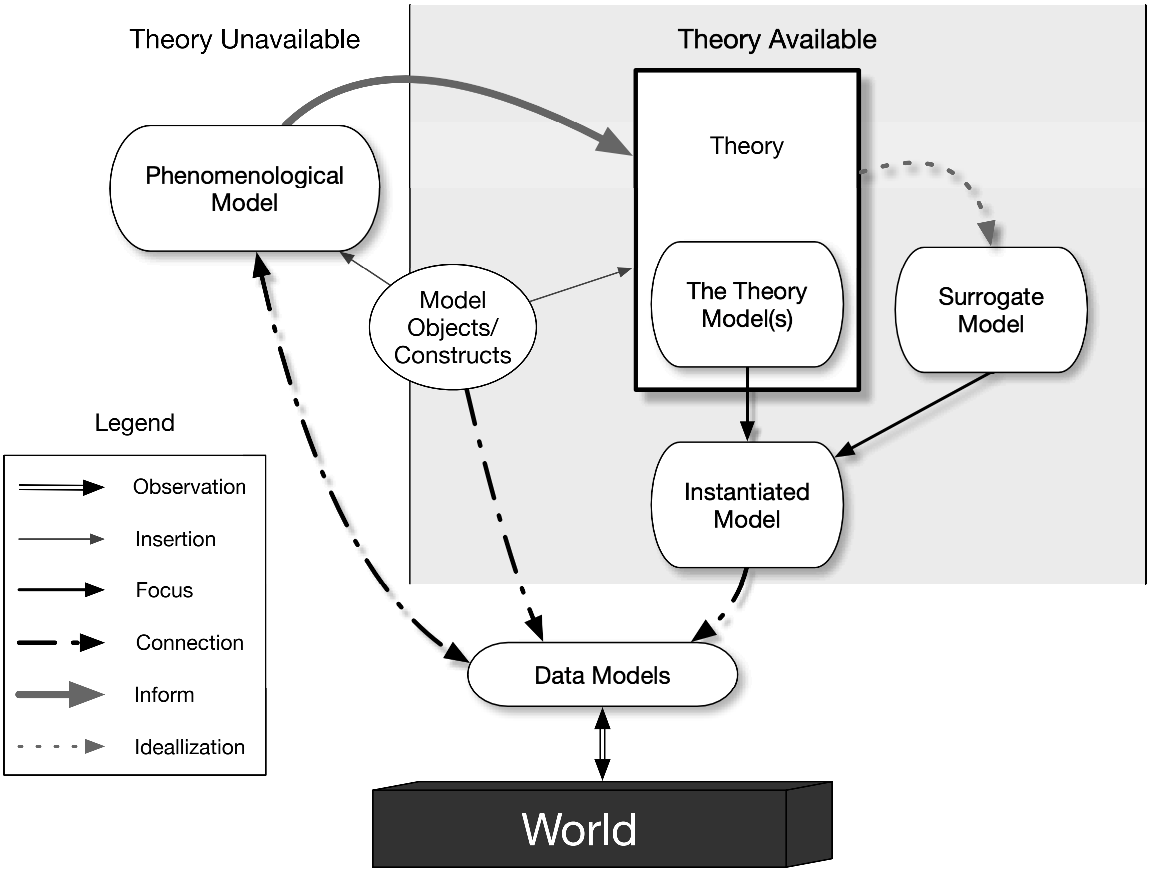


Figure 2.2. The relationship between Theory, model types, and reality. Necessary aspects of a Theory needed for a particular solution are collected and instantiated via focusing, where an instantiated model is created, so it is useful in the context of the specific problem (Boniolo, 2007). If a Theory's mechanisms map to intractable models, convenient models (surrogate models) that do not hold the structure of the theory's mechanism may be substituted through idealization (Weisberg, 2007). Measurement models for the entities operating in the CDs are the objects (constructs) inserted into relational models and Theory. Phenomenological models are constructed from various sources of knowledge in the absence of Theory (Frigg & Hartmann, 2006). These can inform theory development. All models can be directly connected to empirical observation through data models. This figure was adapted from Boniolo (Boniolo, 2007).

Theory. Subsequent expositions of expanded families of related ecological models provided clues as to how inappropriate the models could be as intervention instruments. The early models have been expanded to incorporate stochasticity and spatial dynamics. As an example, when models predicting competitive exclusion (a core concept in ecology) incorporated these extensions, coexistence of competing species became a stable outcome (Lehman & Tilman, 1997). To the extent that the models teach us about the world, this result changes our basic understanding. Quantitative prediction was also affected by the extensions. For example, predicted incidence rates substantially change in infectious disease models when stochasticity is introduced (Renshaw, 1993). These provide important lessons for HP. Models vary widely in their veridicality and utility value, for each specific application (Levins, 1966; Boniolo, 1997; Harré, 2004). This should make us question our confidence in applying models beyond their narrow design purpose. This is an important way models differ from Theory.

There had been some misunderstanding even within ecology about the role of Theory in planning and management applications of mathematical ecology because it was not recognized that: 1) much of theoretical work did not reach beyond modeling, and 2) modeling can progress without a connection to Theory (Kolasa, 2011; Yodzis, 1989). For example, the stochastic and spatial amendments to the population models mentioned above were not an attempt to salvage a tarnished Theory. Scientists can respond to inconsistencies between Theory and observation by adjusting components of Theory that exist in its protective belt (Lakatos, 1968). That is not what was happening in ecology. These modifications did not rescue the hard core of a Theory. There was no Theory. Rather, ecologists have actively collected related bits of understanding regarding how models' behaviors change after amendments, with the hopes that it will provide direction for further investigation. To drive the point home, it was well understood that many ecology models were investigated without any empirical grounding (May, 1981). The point for HP researchers is that ecology has had a robust industry referred to as "theoretical ecology" that has been dominated by modeling and frequently untethered from Theory. If HP does not develop well-formed Engrained Theories it can still progress.

Skeptical Ecologists: Some Models are Dangerous

Ecologists had been divided into two camps (Haller, 2014). One focused on creating and analyzing mathematical models. The other conducted empirical work and harbored some skepticism of the modeling (MacArthur, 1962; McIntosh, 1986). It was considered “fantasy” to believe that the behaviors observed in foundational deterministic ecological models might be observed in the field, where environmental noise destroys any connection between the model and reality (Renshaw, 1993, p. 4). The other side acknowledged that much of the research churn was based on unrealistically simplified models, while noting that mathematical analysis of the properties of sleek models is greatly useful in supporting the understanding of more nuanced models (May, 1981). Even modeling thought leaders warned that reliance on models as a basis of biological knowledge about topics under investigation would be problematic (Cody & Diamond, 1975). Because the models were not derived from Engrained Theory, observational and experimental work was seen as the source of knowledge. In many cases, the models were simply tools to help leverage other sources of knowledge or to help guide the research process. The simplified models of mathematical ecology might appear to present key building blocks of general Theory, until it is recognized that the models, “simply do not hold” (Slobodkin, 1965, p. 348). The models were not intended to be monuments of encapsulated knowledge (Romesburg, 1981). This view of models diverges from the one held in HP.

Ecologists understood there would be negative implications if models that hold little truth value are confused for Engrained Theory. This was demonstrated by warnings that an irrelevant literature from the mathematical ecology industry would accumulate if its work was not better linked to empirical observation (Romesburg, 1981; Werner & Mittelbach, 1981). Ecology's history reminds us that not all CDs have the same truth, or instrumental, value. If HP practitioners seek the benefits of engrained Theory and instead apply a CD with dubious veridicality, disappointment should not be a surprise.

Ecologists' skepticism about obtaining relevant Theories does not arise from a lack of interest in Theory (Haller, 2014). The skepticism results from ecologists' pallet of CDs' inability to function as Theories. The lessons for HP are: 1) a gap between the perceived promises of Theory and what individual CDs actually provide can catalyze skepticism, and 2) phenomenological modeling has benefits, just not the same as Engrained Theory. The expectation gap can be narrowed by presenting benefits as being contingent on the type and status of the CD being used and promoting only strong models for application to health interventions.

We now apply this view to the TPB. Although, deficient and tarnished, when construed as a Theory, it can provide value as a model. As a model it needs only to capture, extend, and translate certain aspects of our knowledge of the about a phenomenon (Harré, 2004). It can be used to investigate part of a system in which the researcher is interested (Minsky, 1965). We know models are wrong to some degree; models are fictive representations of phenomena, with pragmatic value (Boniolo, 2007). A model’s quality depends on its ability to return the relevant properties of the system under investigation to the particular study aims. In the absence of an Engrained Theory to constraint the structure of the model, a phenomenological model can be employed. Models should be chosen and modified by the researcher so that they are optimized for each individual application (see Levins, 1966). This is consistent with recommendations to expertly modify and amalgamate CDs in HP (Glanz & Bishop, 2010). Where health behavior researchers are specifically interested in investigating cognitions and intention in isolation, the TPB can be applied as a phenomenological model. It should be tailored where scientific judgment supports modification or addition of constructs to address specific questions and contexts (e.g. Fishbein & Yzer, 2003).

Dismissed Need and Circumspect Hope: Progress without Theory

It is iconoclastic to question the role of theory in HP, whereas skepticism in ecology was open and multifaceted. Not only was the conflation of models and Theory a skepticism generator, the need for Theory itself was even questioned. Rather than being seen as spurring useful empirical work, some considered theoretical work a distraction. Dayton wrote, “Ecology often seems dominated by theoretical bandwagons driven by charismatic mathematicians, lost to the realization that good ecology rests on a foundation of natural history and progresses by use of proper scientific methods” (Dayton, 1980, p. 156). Theory was not required for advancing knowledge. Although Margalef (Margalef, 1963) stated that ecology was extremely poor in unifying principles, progress and application has been possible, as demonstrated through experiment-informed interventions and development of intervention principles (Braysher, Buckmaster, Saunders, & Krebs, 2012). Yes, the field lacked the blueprint for the core features structuring ecosystems. Prediction and control were also costlier without Theory. Cohesion in the field was limited. Progress was still possible.

Consider how abductive inquiries involving questions about how plant communities are structured led to empirical evidence of the importance cross-species facilitation of population growth (Bruno, Stachowicz, & Bertness, 2003; Callaway, 1997). This type of cross-species interaction was absent in early theoretical work, which focused on competition and predation. Further, ecologists accept that generalizations can be valuable even if they may not lead to the development of deductive Theory (Kolasa, 2011). An example is Holdridge's schematic representation of the association between global vegetation types and dimensions of precipitation and evaporative processes in plant communities (Kolasa, 2011). This work represents important ecological knowledge without the full elucidation of mechanistic processes. Even though it can be argued that ecology can progress without a matrix of well-formed scientific theories, this does not mean the benefits of Theory are discounted (Grime, 2007; Scheiner & Willig, 2011a).

Perhaps, Theory cannot always be obtained. This may be a result of the complexity of the systems under investigation. Ecologists typically study middle-number systems (Allen & Starr, 2017). Its elements and influences are too numerous to enumerate in an idiographic approach, but contextual influences often dominate the system (Pickett, Kolasa, & Jones, 2010). Relative to the scales of determinants, its elements are too few to result in smoothed average signals to investigate. Occurrence of equilibrium states may be rare, as determinants are heterogeneous and unstable. The effect of heterogeneous histories on units of analysis may be the strongest signal on the outcome of interest. Ecosystems exist in unique states in ways that affect future states (Jørgensen et al., 2011). Ecosystems are heterogeneous systems with an immense number of non-exchangeable interacting components (Jørgensen et al., 2011). The slew of challenges for developing explanations and predictions in ecology is further frustrated by investigations of phenomenon that are the result of processes operating at differing temporal and spatial scales (Levin, 1992). HP faces similar challenges. This may explain why Theory is difficult to obtain in both ecology and HP. It might come as a surprise that some areas of physics also share these challenges.

The success of Theory in physics attracts our attention. However, important lessons may be found where the physical sciences operate without a strong Theory. Examples include nuclear physics, which has also been described as a middle-number system (Boniolo, 2007). Practical progress is still achieved in nuclear physics by leveraging principles gained through an abductive approach tied to a patchwork of models (Boniolo, Petrovich, & Pisent, 2002; Boniolo, 2007). Hawking even recognized that some subdisciplines of physics rely on “ad hoc” conceptual devices (Hawking, 2011, p. 112). Relying on phenomenological models that are tailored to the question in hand is similar to how some areas of ecology operate (Weisberg, 2006; Levins, 1966) and should be accepted in HP.

Murry described the barriers to the development of ecological Theory as an attitude that biological systems are too complex, avoidance of imagination initiated deductive-nomothetic approaches, and conflation of CD types (Murray, 2001). Ecologists have had differing views, but there is no disagreement that Theory sparsity slows progress. In his history of ecology, McIntosh concluded that ecology's progress was also slowed by the lack of a consistent, coherent, and lucid conception of Theory (McIntosh, 1986). This was, in part, a result of an ambiguous and imprecise lexicon related to the topic. McIntosh's conclusion is eminently relevant to HP. Fortunately, it is surmountable by adopting a more descriptive vocabulary.

DISCUSSION: HOW DO LESSONS FROM ECOLOGY MAP TO HEALTH PROMOTION?

Ecology's modeling industry did not create disciplinary cohesion. The discipline resembles, “an amorphous, post-modern hotel and rabbit warren with separate entrances, corridors, and rooms that safely accommodate the irreconcilable" (Grime, 2007, p. 227). This state existed in ecology despite the experimental literature commonly invoking theory (Kolasa, 2011; Scheiner, 2013). Ecology and HP are in analogous situations–the underlying issue appears to be a desire to use theory whilst broad, Engrained Theory is scarce. Ecology's history offers multiple strategies that HP can use as a response to this situation (Table 2.3). The first is to inaugurate new traditions of method and conceptualization that organize the field. HP has done this before with its movement toward ecological approaches (Sallis, Owen, & Fisher, 2015). Similarly, ecology has also shown how there can be paradigm adjustments without a new content-based Theory (Chesson & Case, 1986; Kolasa, 2011). For example, hierarchy theory lays out how researchers may approach ecosystem processes operating at different scales, without making specific predictions about any specific ecosystem (Allen & Starr, 2017). One view of ecology is that it may not be able to anchor itself in traditional hypothetico-deductive Theory, as some physical sciences have, but it can still organize itself using conceptual frameworks (Pickett, Kolasa, & Jones, 2010). Where Theory is sparse, integration of various forms of knowledge, modeling, and abductive research programs has been successful in addressing important ecological questions (National Research Council, 1986; Shrader-Frechette, 2008). Accepting the use of Theory-free conceptual frameworks is the first approach that HP can adopt.

A second approach is to build bounding CDs to help guide research. Two examples are the work lead by Scheiner and Willig and the other by Jørgensen (Jørgensen et al., 2011; Scheiner & Willig, 2011b). The bounding CDs are lists of fundamental principles of broad disciplinary scope, which constrain subtending CDs (Scheiner & Willig, 2011a). In this scheme, Theories with narrower scope, such as the Theory of Island Biogeography, are nested as

Table 2.3. Complementary strategies for structuring HP research in the absence of Theory.

Process Domain	Feature	Approach	Notes on Related Literature
Lexical evolution	Differentiation among conceptual devices	Extend the field's vocabulary to increase the salience of the different qualities of theories, models, and frameworks without attempting to reform terms currently in use.	Hodges (2008) recognized that attempts to reform current terminology require consensus and current users to adopt new norms, while future researchers would need to avoid adopting applications found in the legacy literature (Hodges, 2008). Expanding the vocabulary avoids this and facilitates conceptual clarity.
Lexical evolution	Identification and declaration of theory status	Develop a nomenclature to facilitate the consistent description of Theory status.	Recommendations about use of Theory based on Theory status may improve expectations related to Theory. Both Kolasa (2011) and Pickett (2010) outlined the importance of evidentiary status in use of Theory.
Theory development	Bounding principles	Develop suites of bounding principles that can guide conceptual development of models and constituent Theories	Jørgensen et al. (2011) and Scheiner and Willig (2008) present exemplars of bounding principles that can help guide research in sub-disciplines where Theory is sparse.
Research framing	Phenomenological modeling	When Theory is unavailable, build and use models that are untethered from Theory	Boniolo (2007) provided a cogent description of the role of scientific modeling in the absence of Theory.
Research framing	Abductive inference	Allow for some research hypotheses to be generated from the creative integration of observations and analysis.	Haig (2005) described research that iterates through best guesses given the current body of knowledge and updating assessments as new evidence emerges (Haig, 2005).
Research framing	Paratheoretical programs	Allow for some important research questions to be answered in the absence of connections to H-D Theory	Where Engrained Theory is unavailable to apply to an important research question, application of sound scientific methodology can still provide valuable answers and potentially inform future Theory development.

Constituent Theories within the fundamental principles. The constituent CDs must conform to the principles. The intent has been to provide conceptual scaffolding for the whole of ecology, so the organizing and cohesive benefits of Theory can be obtained. To be clear, these general theories are a “list of fundamental principles” and a set of "basic principles," which can be combined to predict bounds on ecosystem processes and patterns (Scheiner & Willig, 2011a, p. 4; Jørgensen and Fath, 2004). These are not archetypal hypothetico-deductive Theories.

These principles are presented in conjunction with other assumptions, definitions, logical or causal structures, and explication of the domain of applicability. Multiple processes are expected to be found within these bounding CDs. It is the constituent CDs that are expected to provide precise structures to be instantiated in individual predictive models. Even laws in ecology would occur at the constituent level (Scheiner & Willig, 2011a). Examples of principles of ecology include: 1) organisms are distributed in space and time in a heterogeneous manner, and 2) birth and death rates are a consequence of interactions with the abiotic and biotic environment (Scheiner & Willig, 2011a). The proponents of these bounding CDs recognize these as basic axioms related to population processes. In isolation, the set of principles do not provide much service to a practitioner. Their careful formal listing helps hone the science and ensure that construction and evaluation of constituent CDs reflect these principles.

Another example is Jørgensen's et al.'s Ecological Law of Thermodynamics (ELT) (Jørgensen et al., 2011). This work was influenced by considerations of thermodynamics and is applicable to ecosystems. Principles form the core of this CD, rather than laws, because strong laws are elusive in ecology, as they are in HP. Ecosystem dynamics are so complex that contextual influences can perpetually disrupt or cloud signals from law-like mechanisms (Abbott, 1980; Jørgensen et al., 2011). Some general boundaries on empirical patterns are predicted by ELT. This CD suggests that greater through-flow of exergy in an ecosystem results in increases in its exergy, moving further away from thermodynamic equilibrium. Where ecosystems are not constrained to a single possible trajectory for the organization of the system components and processes, the configuration that results in the highest exergy under the prevailing conditions will be selected (De Wit, 2005; Jørgensen et al., 2011). This CD is expected to provide a starting point for further knowledge building, which will ultimately aid practitioners (Jørgensen et al., 2011). The proposed benefits of the CD are more subdued than expectations of Theory. Jørgensen et al. wrote, "some predictability is still possible, although we should expect accuracy to be small and uncertainty to be high" (Jørgensen et al., 2011, p. 57). Success of the ELT would demonstrate that establishing the boundary conditions that tie together and constrain constituent CDs can be accomplished in ecology. Jørgensen admits that it "will be much more difficult to develop an applicable, predictive ecological [T]heory" (Jørgensen et al., 2011, p. 167). Thus, even contemporary ecologists at the forefront of theoretical developments are not optimistic about achieving well-formed Theories. Instead, their hope is for unification through CDs that bound subtending research and theory. HP may benefit from taking a similar posture.

Paratheoretical Framing

Important questions and approaches to hypothesis generation can occur without Theory (Dochtermann & Jenkins, 2011). Although explicit prescription of abductive scientific approaches is uncommon in ecology, some ecologists have presented the view that some model building should be iterative and include abductive processes (Burnham & Anderson, 2003, p. 14, 17; Burnham et al., 2011). As in the behavioral sciences, there have been sporadic calls for greater acceptance of abductive processes (Griffin, 2006; Haig, 2005; Muthukrishna & Henrich, 2019). If these alternative processes for framing research are not viewed as acceptable by reviewers, some HP researchers may patch together disparate and distant Theories to be invoked as justification for hypotheses or interpretations. This risks misleading readers regarding: 1) the authentic conceptual underpinnings of the research, and 2) the relevancy of the empirical findings to these distant Theories. We suggest an alternative that can be described as paratheoretical inquiry. This approach is abductive and exploratory. A paratheoretical process integrates empirical observation and modeling through iterative development of explanations. Importantly, it does not require Theory to operate, yet connects to Theory when opportunities for material contributions to theoretical development arise.

Incorporating Alternative Approaches

In summary, ecology and HP face similar challenges. Both are sparsely covered by Engrained Theories. Both have histories of internal factions advocating for theory. A consequence of the heterogeneous complex systems that both investigate is the difficulty in obtaining sleek predictive CDs. Theory has been identified as the device pivotal for achieving desired prediction and control in both fields. Indeed, both have a history of theory being considered too disconnected from practice. Although they have shared challenges, the fields now hold different postures toward theory.

Ecology's theoretical work has been focused on constructing and analyzing models. Ecologists understand that all models are wrong, and some have no proximal relevance to empirical reality. It would be malpractice for practitioners to use some models as a basis of their interventions. On the other hand, Engrained Theory can be treated as adequately correct and the superior option for constructing interventions. Although some blurring has occurred, ecologists recognize the distinction between Theory and models. There is also a strong recognition that models vary in their structure and utility and that their predictive capacities vary widely. HP has taken a different route in its relationship to theoretical investigations. Health behavior researchers have not consistently recognized or applied a distinction between Theories and models (Bartholomew & Mullen, 2011; Lippke & Ziegelmann, 2008; Nilsen, 2015). Recommendations to use theory are devoid of clarifications about the type of CD to be used. It leaves the field vulnerable to the belief that any CD can be used interchangeably, or in ad hoc combinations, and still have met the field's expectation for leveraging theory.

If HP researchers choose to follow ecologists, a first step would be the expansion of their conceptual and lexical approach to theory. This step is likely to result in the acceptance that strong Theory may be elusive and unnecessary for progress in HP. HP is pre-positioned to strengthen its commitment to abductive approaches. Rothman's suggestion for iterating HP relevant models through specification, application, evaluation, and refinement to obtain adequate *models* is feasible (Homer, 1996; Rothman, 2004). Lexical expansion, in combination with a disciplinary focus on modeling, will help resolve discrepancies between expectations and outcomes from the use of theory. Closing this gap may reduce the propensity for practitioner skepticism and bring practitioners into the fold, as their role should be fundamental to adaptive model building. A combination of bounding principles and adaptive model building in paratheoretical research settings would be useful in guiding research in both established and expanding areas of inquiry.

ACKNOWLEDGEMENTS

The author of this research would like to acknowledge all the participants for their participation in the study. I am also grateful for Professor Miguel Acevedo's thoughtful and engaging discussions regarding contemporary views on the use of theory in ecology.

Chapter 2 has been submitted for publication and is under review as: Cummins, K. M., Zuniga, M. L., Reed, M., Pitpitan, E. V., & Brown, S. A. Reforming the use of theory in health promotion: lessons from ecology. *Health Psychology Reviews*. Kevin Cummins was the primary investigator and lead author of this paper. This paper was co-authored by María Luisa Zúñiga, Mark Reed, Sandra Brown and Eileen Pitpitan.

CHAPTER 3: COMPARISON OF FACTOR ANALYSIS MODELS APPLIED TO THE NCANDA NEUROPSYCHOLOGICAL TEST BATTERY

ABSTRACT

The factor structure of neuropsychological functioning among a large sample (N = 831) of American youth (ages 12-21 at baseline) was investigated. Candidate models were selected based on their potential to provide service to the study of adolescent development and the effects of heavy episodic alcohol consumption. Data on neuropsychological functioning were obtained from the NCANDA study. This is a longitudinal community study of the effects of alcohol exposure on neurodevelopment. Three conceptually motivated and one empirically motivated factor analysis model of neuropsychological domains are compared based on penalized-likelihood selection criteria and model fit statistics. Two conceptually-motivated models were found to have adequate fit and pattern invariance to function as a measurement model for the WebCNP-anchored neuropsychological battery in NCANDA. Corroboration of previous factor analysis models was obtained, in addition to the identification of an alternative factor model that has higher discriminant capacity for neuropsychological domains hypothesized to be most sensitive to alcohol exposure in human adolescents. The findings support the use of a factor model developed originally for the WebCNP and a model developed specifically for the NCANDA project. The NCANDA model 8-Factor Model has conceptual and empirical advantages that were identified in the current and prior studies. These advantages are particularly valuable when applied in alcohol research settings, because cognitive functions known to be sensitive to alcohol's effects are captured by this model.

INTRODUCTION

Heavy alcohol consumption can result in brain damage. Prolonged exposure to high levels of alcohol, as experienced by some adults with alcohol use disorder, can result in brain tissue shrinkage or even atrophy as a consequence of behaviorally mediated and direct neurotoxic mechanisms (Oscar-Berman & Marinkovic, 2003; Oscar-Berman et al., 2014). Noted brain regions that are susceptible include areas involved in learning and memory (basal forebrain and limbic system), posture and motor control (cerebellum) and, particularly, executive functioning (prefrontal cortex) (Sullivan et al., 2000; Le Berre, 2019). White matter, the network of tracts allowing inter-regional communication, is also broadly vulnerable to the chronic effects of alcohol exposure in humans (Zahr & Pfefferbaum, 2017). The neurocognitive consequences of heavy alcohol exposure can include memory, learning, inhibition and executive functioning deficits, in addition to emotional blunting (Ryan & Butters, 1980; Le Berre, Fama, & Sullivan, 2017). However, these findings were born from the study of adults with years of excessive alcohol exposure and may differ in degree and features from the types of effects experienced by community adolescents, where typical drinking patterns are intermittent.

Adolescence is an important period of neuromaturation and a sensitive period for persistent alterations to the brain's reward circuitry resulting from alcohol exposure in animal models (Crews et al., 2016; Salling et al., 2018; Spear, 2018). In humans, drinking onset during this period is a known risk factor for alcohol use disorder(Grant et al., 2006; Grant & Dawson, 1997). Emerging evidence suggests that patterns of alcohol exposure experienced by community youth can result in alterations in neuroanatomical development (Squeglia et al., 2009b). Neuropsychological deficits associated with adolescent alcohol exposure include verbal learning (Mahmood et al., 2010; Sneider et al., 2013; Nguyen-Louie et al., 2016), visuospatial processing

(Tapert & Brown, 1999), executive functioning (Mota et al., 2013), attention (Tapert & Brown, 1999; Tapert et al., 2002) and memory (Mota et al., 2013; Hanson et al., 2011; Tapert et al., 2002; Nguyen-Louie et al., 2016). The functional implications of alcohol-induced neuromodulation are just now being confirmed in large-scale longitudinal community samples (Brown et al., 2015; Jernigan et al., 2018; Akshoomoff et al., 2014; Anonymous, 2014). These studies are designed to provide a more definitive evaluation of adolescence as a sensitive period for the neurological influence of substance use exposure in humans.

The National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) is one of the large-scale studies designed to evaluate neuropsychological changes associated with alcohol use. To address its objectives, NCANDA's protocol includes a neuropsychological test battery specifically designed to be sensitive to hypothesized decrements in neuropsychological functioning due to alcohol exposure. Although a description of the cross-sectional patterns observed among the conceptually-driven composite scores of domain-specific tests has been reported in NCANDA (Sullivan et al., 2016), evaluation of a data-driven factor structure in the NCANDA sample has not been reported.

The application of factor analysis to neuropsychological test batteries has been common. Accounts of human intelligence and cognitive architecture, for example, Halstead's (1947), Newby et al.'s (1983), and Patt et al.'s (2017), are anchored in factor analysis. In addition to providing insight into cognitive architecture, factor analysis has been employed to achieve the statistical benefits of dimension reduction and reduced estimation error (Genderson et al., 2007). Substantial interest in the use of factor analysis has been generated in the study of changes in cognitive function associated with neurodevelopment and psychopathology (Moleiro et al., 2013; Thompson et al., 2019; Patt et al., 2017; Masterson, Tuttle, & Maerlender, 2019). Factor analysis has also been applied to the Web-based Penn Computerized Neuropsychological Battery (WebCNP; PennCNB)(Moore, Reise, Gur, Hakonarson, & Gur, 2015), which comprises the core components of the NCANDA test battery (Sullivan et al., 2016). When applied to a general population of youths, the factor structure of the WebCNP tests was found to be similar to the conceptual organization (Gur Model) used to develop this standardized battery (Moore et al., 2015). The difference between the conceptually motivated model and the empirically based model was that the WebCNP's conditional exclusion test loaded onto the Complex Cognition latent factor in the empirically motivated (bifactor) model based on efficiency scores, as opposed to the Executive Functioning factor, when constructed as a correlated-traits model (Moore et al., 2015). The conditional exclusion test was designed to measure abstraction and mental flexibility. An alternative conceptually justified organization of the NCANDA test battery, which includes the WebCNP, was developed in Sullivan et al. (2016). Application of this organization breaks executive functioning into attention and working memory (Table 2.1).

The present analysis aimed to replicate prior findings regarding the factor structure of the WebCNP (Moore et al., 2015) and extend these findings by testing model invariance. A second novel aim was to identify the optimal factor structure for the NCANDA neuropsychological battery from a set of candidate models, which includes the Gur Model and empirical (data-driven) factor models developed specifically for the WebCNP, and a configural model based on Sullivan et al.'s (2016) neuropsychological-informed conceptual analysis (8-Factor Model). A final aim was to evaluate a new conceptually driven configural model motivated by an intent to maximally tap the neuropsychological functions reported to be the most vulnerable to alcohol exposure in human adolescents. To that end, a Vulnerabilities Model was investigated that included factors designed to tap visuospatial processing, attention, memory, and verbal learning,

Table 3.1. Complementary strategies for structuring HP research in the absence of Theory.

Assessment		Domain	Suite	8-Factor	Vulnerabilities†	Gur+
	References			Sullivan et al., 2016	Spear, 2018	Moore et al., 2015
Continuous Performance	Gur et al., 2010	Attention	1,2	Attention	Attention	Executive Functioning
N-Back Task	Gur et al., 2010	Working Memory	1,2	Working Memory	Attention	Executive Functioning
Penn Face Memory	Gur et al., 2010	Face Memory	1,2	Episodic Memory	Memory	Episodic Memory
Penn Delayed Face Memory		Face Memory	1,2	Episodic Memory	Memory	Episodic Memory
Visual Object Learning	Gur et al., 2010	Spatial Memory	1,2	Episodic Memory	Memory	Episodic Memory
Delayed Visual Learning		Spatial Memory	1,2	Episodic Memory	Memory	Episodic Memory
Penn Word Memory	Gur et al., 2010	Verbal Memory	1,2	Episodic Memory	Verbal Learning	Episodic Memory
Continuous Exclusion	Gur et al., 2010	Mental Flexibility	1,2	Abstraction	Visuospatial Processing	Executive Functioning
Penn Verbal Reasoning	Gur et al., 2010	Language Reasoning	1,2	Abstraction	Verbal Learning	Complex Cognition
Matrix Reasoning Test (A)*	Gur et al., 2010	Nonverbal Reasoning	1,2*	Abstraction	Visuospatial Processing	Complex Cognition
Emotion Differentiation	Gur et al., 2010	Emotion Differentiation	1,2	Social Cognition	Social Cognition	Social Cognition
Emotion Recognition	Gur et al., 2010	Emotion Identification	1,2	Social Cognition	Social Cognition	Social Cognition
WRAT-4: Math	Wilkinson et al., 2006	Math Computation	1	General Ability	General Ability	*General Ability*
WRAT-4: Reading	Wilkinson et al., 2006	Word Reading	1	General Ability	General Ability	*General Ability*
Vocabulary	Sullivan et al., 2016	Vocabulary	1,2	General Ability	Verbal Learning	*General Ability*
WAIS-4: Digit Symbol	Wechsler, 2008	Complex Attention	1	Motor Speed	Motor Speed	*Motor Speed*
Grooved Pegboard	Matthews & Kløve, 1964	Dexterity	1	Motor Speed	Motor Speed	*Motor Speed*
Walk-a-Line	Fregly et al., 1972	Ataxia	1	Balance	Balance	*Balance*

Note. Italics indicate supplemental tests. Shared colors indicated shared factor structures.

which are neuropsychological functions believed to be most sensitive to alcohol exposure in adolescence (Spear, 2018). Establishing a measurement model for these vulnerabilities could enhance the estimation of associations between target neuropsychological constructs and alcohol consumption. Accordingly, our overarching aim was to identify which factor structures are applicable to contemporary measurement batteries in youth samples in the context of substance use research.

METHODS

Sample

Data originate from the NCANDA project. NCANDA is a longitudinal study of youth that employs a modified accelerated longitudinal design (Miyazaki & Raudenbush, 2000; Galbraith, Bowden, & Mander, 2017; Brown et al., 2015). Recruitment and enrollment ($n_{\text{participants}} = 831$, $n_{\text{families}} = 670$) resulted in samples demographically representative of each metropolitan catchment region ($n_{\text{regions}} = 5$), with enrichment of participants at risk for lifetime substance use disorder; 51% of the participants reported at least one of the following: family history of alcohol use disorder; externalizing or internalizing symptoms; or consumption of alcohol before age 15. Exclusionary criteria included magnetic resonance imaging contraindications (e.g., permanent metal in the head such as dental braces), neurotropic medications, serious medical problems, major Axis I disorders, pervasive development disorder, uncorrectable vision or hearing impairment, lack of English fluency, and substance use disorder. As details of the study design were previously published (Brown et al., 2015) only a sketch of the protocol is presented here. Extensive quantitative descriptions of the neuropsychological scores can be found in Sullivan et al. (2016). At project baseline, participants were 12-21 years old, with oversampling of younger ages and those with no more than limited substance use experience. Participants' neuropsychological functioning was assessed annually. The protocol was reviewed and approved by the human research protection program at each participating university.

Measures

Neuropsychological functioning was assessed with the WebCNP (Gur et al., 2010), plus several traditional tests including the Wide Range Achievement Test-4 (WRAT4), Word Reading and Arithmetic sub-tests (Wechsler, 2008), Grooved Pegboard Test (Lezak, Howieson, Loring, & Fischer, 2004), and Fregly-Graybiel Walk-a-Line (FGWL) postural stability test (Graybiel et al., 1972; Sullivan et al., 2000). The pegboard tests manual dexterity with a timed score for the completion of a peg insertion task with each hand. The FGWL assesses ataxia. WebCNP assesses a range of neuropsychological domains and provided 12 separate and composite test scores based on performance accuracy and speed, spanning various domains of functioning (Table 3.1). The WebCNP battery was slightly modified from the standard WebCNP (a.k.a. PennCNB) distribution to optimize the sensitivity of the battery to detect the effects of alcohol exposure. To better distinguish immediate-recall and delayed-recall (Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003), delayed versions of Visual Object Learning and Penn Face Memory tests were substituted for the Penn Line Orientation Task and the Age Differentiation Task. The immediate and delayed versions of the memory tests were found to have strongly correlated errors and were combined (meaned) for this study. Accuracy scores, which are available for the WebCNP and the traditional tests, were used in this study.

Demographics, including sex and age, parental socio-economic status (income, occupation and educational attainment), were obtained during the baseline interview via self-report as described in Brown et al. (2015). Familial relationships among participants were also identified through proband and parental self-report.

Statistical Analysis

Confirmatory Factor Analysis

Using confirmatory factor analysis (CFA), models were fit to the baseline NCANDA data. The models were fit to the data in the following sequence. Configural CFA models were fit using maximum likelihood. Targeted models that failed to converge were reestimated with starting values from pooled models, with Newton-Raphson iterations, and run using Stata's *difficult* algorithm. The reported loadings are based on standardized solutions of the CFA models. Model configurations are presented in Table 3.1. Neuropsychological scores were treated as reflexive indicators (Bollen & Bauldry, 2011). To account for the non-independence of participants within families, models were initially fit (using Stata's *gsem*) with participants nested within families (Wu, Lin, Nian, & Hsiao, 2017; Huang & Cornell, 2016). This structure resulted in the failure of most models to converge, even with substantial mitigation. In response, the modeling approach was modified. Models were run in Stata 15.1 using the *sem* functions (StataCorp, 2017). The model comparison results were based on bootstrap estimates, where each bootstrap sample consisted of only one participant from each family. Means of up to 1,000 bootstrap samples are reported. The final models used in each bootstrap sample estimation were allowed to run out to 100 iterations. If convergence was not achieved, the statistics were treated as missing. In models that included single indicators on latent constructs, the variances of the indicator(s) were constrained to a constant equal to one minus the reliability (Brown, 2015). Reliability estimates were extracted from prior reports (Gur et al., 2010; Fregly, Graybiel, & Smith, 1972) and unpublished data (Moore, 2019). Latent variables with single indicators do not provide the measurement benefits resulting from the use of measurement models but can provide utility to the overall substantive investigation (Hayduk & Littvay, 2012). *A priori* loadings were available to be used in metric models for the Empirical Model (Moore et al., 2015). Because the version of the WebCNP used in the NCANDA study included two unconventional tests, these were left free in all metric models. All other parameters, including loading for traditional neuropsychological tests, were left unconstrained. All first order models were correlated traits models.

At baseline, 99% (n_{WebCNP} = 828) of participants completed the full WebCNP assessment battery, and 97% ($n_{\text{NCANDA Battery}}$ = 806) completed all neuropsychological assessments. Missing data was deleted listwise (i.e. complete case analysis was employed), because the most common missing value pattern was missing the entire neuropsychological battery (75% of cases with any missing WebCNP tests were missing all the tests). This pattern limits the benefit of missing value analysis, where there are no auxiliary variables available that can provide proxy information on the missing observations. Eleven participants were missing both the ataxia and math ability assessments. The employed analyses assume missing values are missing completely at random (Little & Rubin, 1986).

Data were assessed for normality before model estimation. Univariate distributions were evaluated through visualization with histograms and Q-Q plots, and estimation of higher order moments. Multivariate distributions were assessed via scatterplot matrices, estimation of Mahalanobis distances, and evaluation with the BACON algorithm, set to detected outliers beyond the 15th percentile of the $\chi^?$ distribution (Billor, Hadi, & Velleman, 2000; Weber, 2010). As no overt two-way curvilinear associations or grossly non-normal distributions were identified, data were left untransformed.

Model Fit Evaluation

Evaluation of model fit was conducted by gauging fit indices against conventions outlined by Hu and Bentler (Hu & Bentler, 1999a). Bright-line application of conventional cut-points is avoided in recognition of the graded nature of model adequacy (Marsh, Hau, & Wen, 2004). Further, the focus was made on comparative fits among models rather than the fit of individual models (Marsh et al., 2009). We used likelihood ratio tests to evaluate the level of significance between the target model and the saturated and base models, under a neoFisherian evaluation framework (Hurlbert, Levine, & Utts, 2019; Cummins & Marks, 2020). Incremental fit was assessed with the comparative fit index (CFI). Absolute fit was assessed with root mean square error approximation (RMSEA) and the standardized root mean square residual (SRMR). Values for these were viewed in light of conventional cut-offs of CFI > 0.95, RMSEA < 0.05, and SRMR ≤ 0.08 for identifying well-fitting models (Hu & Bentler, 1999a; Browne & Cudeck, 1993). Models were deemed clearly unacceptable if CFI < 0.80, RMSEA > 0.10, and SRMR > 0.10 (Hu & Bentler, 1999a). Although there are additional fit indices that can be used in model evaluation, we present a limited preselected set of indices with adequate statistical properties, coverage, and interpretability.

Model Comparison

Two nested sets of indicators were evaluated. The broadest group used all of the available neuropsychological tests for the NCANDA sample, this is referred to as the NCANDA suite (16 indicators). The conceptually motivated models were applied to this set as configural models (Table 3.1). Because the Gur Model did not originally incorporate all of the NCANDA assessments, it was extended by adopting the conceptual structure outlined by Sullivan et al. (2016) for the additional assessments that are included in the full NCANDA suite. The extended model is denoted by a plus. Because an aim of this paper, and the focus of prior empirical research, has been on the factor structure of the WebCNP tests, a suite of models limited to the WebCNP battery are also evaluated (10 indicators; see Table 3.1). All of the conceptual models are applied to both assessment suites. An empirically developed model was also available to provide a benchmark for comparisons. The empirically motivated 3-factor model (Empirical Model) was based on an exploratory factor analysis model of WebCNP accuracy scores reported by Moore et al. (2015). Because loadings were available from the earlier report they were used to construct a metric version (constrained loadings only) of the Empirical Model (Moore et al., 2015). There was no empirical basis to extend the Empirical Model to cover the full NCANDA assessment suite, so this was estimated only in the WebCNP suite.

The Empirical Model was included as a supplement to the conceptually grounded set of models. The decision to supplement the set of models was made after an initial investigation of the preplanned models. In accordance with recommended reporting practices, all of the post-hoc analyses were identified and interpreted as exploratory research steps (Appelbaum et al., 2018; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

Model comparisons were primarily based on a set of penalized-likelihood selection criteria (AIC, BIC)(Kuha, 2004), and supplemented by CFI, RMSEA, and SRMR (Sellbom & Tellegen, 2019; Raftery, 1995; Burnham & Anderson, 2003; Schmitt, Sass, Chappelle, & Thompson, 2018). Where discrepancies between AIC and BIC model selection occurred, BIC selection prevailed because it is more consistent with our objective of approximating the correct model and giving deference to parsimony, rather than optimizing a predictive model (Burnham, Anderson, & Huyvaert, 2011; Burnham & Anderson, 2003; Aho, Derryberry, & Peterson, 2014).

Invariance Testing

Once optimal models were identified from among the candidate CFA models, configural invariance was evaluated. For this purpose, three series of multi-group CFA models with metric and scalar parameter constraints were fit for age groups (< 16.5 years old, ≥ 16.5 years old), self-identified sex, and measurement waves (baseline through year 4). Although configural model fit adequacy was considered necessary, it was not treated as sufficient (Hayduk, 2014). Having the same pattern of salient factor loadings across groups was considered supportive of a configural invariance finding (Steenkamp & Baumgartner, 1998). After configural invariance was ascertained, multiple-group CFAs were estimated. Evaluation of metric invariance began by estimating a model where all parameters were freely estimated. This was defined as the configural model, which was then compared with a model that included equality constraints on the loadings across groups (the metric model). A scalar model with constrained loadings and intercepts was next estimated. Intercepts for one of the indicators of each factor was constrained to zero to address identifiability in the group-CFA models.

Likelihood ratio tests comparing these nested models were used to gauge the significance of the added restrictions in each successive model (Kim & Willson, 2014). However, this mode of evaluation is confounded by sample size (Putnick & Bornstein, 2016); it does not separately identify features or gauge the magnitude of misconfiguration. Thus, differences in the parameter estimates were also inspected to evaluate their contributions to the differences among groups. Further, an additional set of model fit statistics were inspected to gauge the change in model fit as a consequence of each successive set of constraints. Differences in CFI and RMSEA were evaluated in light of recommendations that ΔCFI and ΔRMSEA should not exceed -0.010 and 0.015, respectively (Chen, 2007; Chen, 2007; Cheung & Rensvold, 2002; Little, 2013)(cf.

Johnson and Braddy 2008). Deference was given to ΔCFI, as recommended by Sellbom and Tellegen (2019). Where measurement invariance was not achieved, model modification indices (MI) were used to assist in the identifications of parameters that may be contributing to worsening model fit (Byrne, Shavelson, and Muthen 1989). Joint tests of modification indices (MI) were applied in each model, first with Wald tests of all free parameters, evaluating if they significantly vary across groups, then with *score* tests of parameters constrained to equality across the groups (see Harrell [2015] for a description of these classes of hypothesis tests). Parameters in the joint tests were limited to the loadings in the configural and metric models and included intercepts in the scalar model. Partial invariance was investigated by refitting the model with parameters identified by the MI being freely estimated, without constraint. Interpretation of invariance violations among groups relied on inspection of loading patterns in each subset.

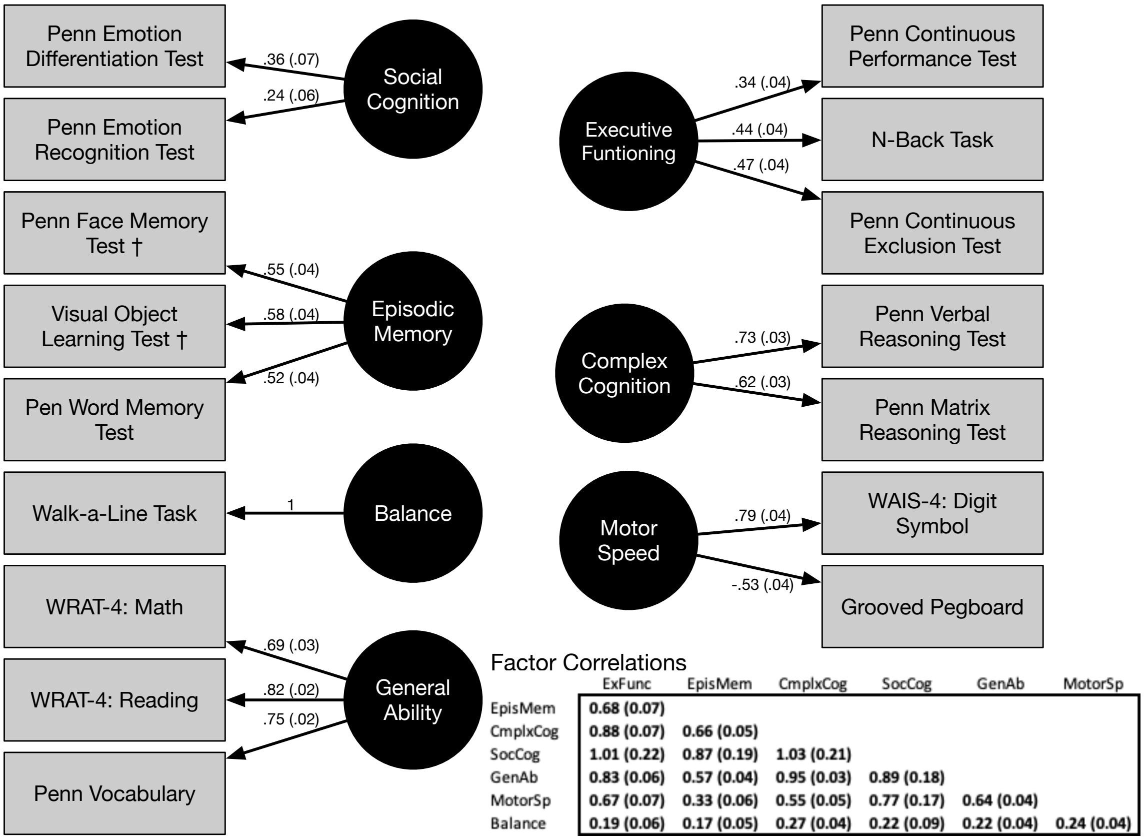
RESULTS

Sample Characteristics

At baseline, ages ranged from 12.1 to 22.0. The median was 15.9 years with the 24th and 75th percentile for age at 14.1 and 18.0. Boys accounted for 49% of the sample. Most participants represented the 670 families in the NCANDA sample as singletons ($n_{\text{singletons}} = 531$) with 17 families contributing more than two children to the sample ($n_{\text{two siblings}} = 244$, $n_{> 2 \text{ siblings}} = 56$). Of all families, 47% had at least one parent with a post-baccalaureate degree. Most of the remaining families had a parent with an undergraduate degree (43% of families). The highest degree was a high school diploma or equivalent for 8% of families; 1% of families had parents without a diploma. At baseline, the mean number of days participants had used alcohol and cannabis during their lifetime was 9.2 (SD = 35.4) and 8.6 (SD = 85.3), respectively. See Brown and colleagues (Brown et al., 2015) for a detailed description of the sample.

Model Comparisons

Application of the conceptually-motivated CFA models under investigation in this study resulted in the identification of multiple models that had a constellation of moderately strong fit statistics when applied to the WebCNP suite of neuropsychological tests. However, when applied to the full NCANDA suite, the fits were only marginally sufficient. Application of BIC as the model selection criterion, in this broader suite, identified the Gur+ Model as the superior model (Table 3.2). The RMSEA point estimate for the Gur+ Model was at 0.06 with an associated probability of the population value being under 0.05 that was low ($P_{close} > 0.05$). The baseline comparison statistic for this model was moderate; CFIs were 0.92. The residual magnitude



	ExFunc	EpisMem	CmplxCog	SocCog	GenAb	MotorSp
EpisMem	0.68 (0.07)					
CmplxCog	0.88 (0.07)	0.66 (0.05)				
SocCog	1.01 (0.22)	0.87 (0.19)	1.03 (0.21)			
GenAb	0.83 (0.06)	0.57 (0.04)	0.95 (0.03)	0.89 (0.18)		
MotorSp	0.67 (0.07)	0.33 (0.06)	0.55 (0.05)	0.77 (0.17)	0.64 (0.04)	
Balance	0.19 (0.06)	0.17 (0.05)	0.27 (0.04)	0.22 (0.09)	0.22 (0.04)	0.24 (0.04)

Figure 3.1. Confirmatory factor analysis estimates of the Gur+ Model for the NCANDA suite of tests.

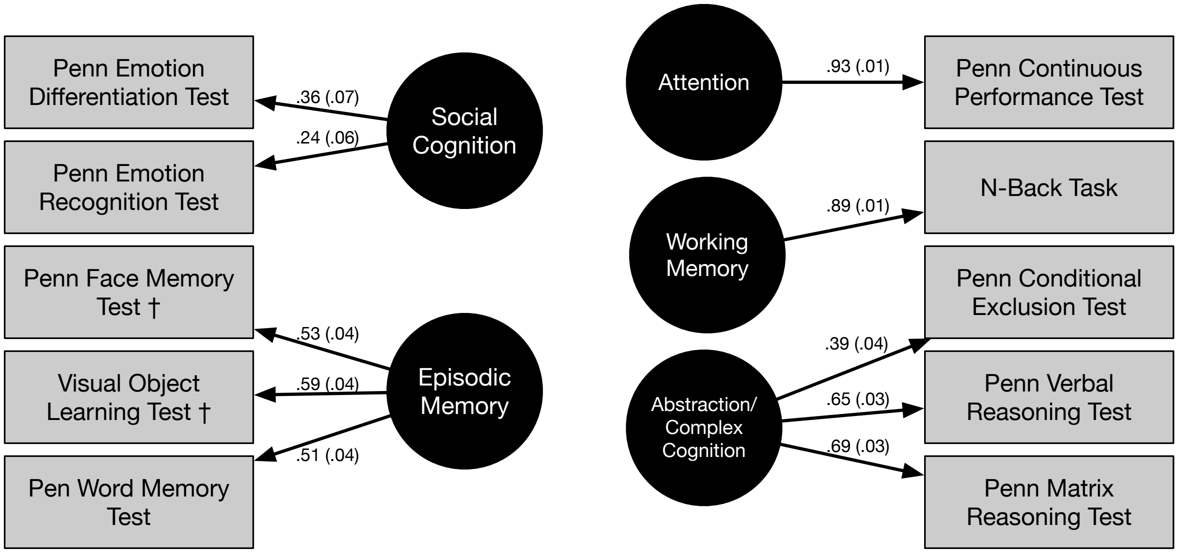
statistic evidenced adequate fit; the model had an SRMR of 0.04. The model fit of the competitors was similar to the Gur+ Model (Table 3.2). In addition, the 8-Factor Model's BIC was of similar magnitude (0.8 points higher), which is an equivocal difference based on a 5 point rule for comparison (Anderson & Burnham, 2004). Further, the 8-Factor Model had a substantially superior AIC (21.6 points lower).

The strongest loadings in the Gur+ Model were found with the general ability, motor speed, and complex cognition latent factors ($|\boldsymbol{\lambda}|$'s > 0.50, Figure 3.1). Episodic memory indictors were moderately loaded ($\boldsymbol{\lambda}$'s of 0.52 to 0.58). Executive Function loadings were lower,

Table 3.2. Confirmatory factor analysis model fit comparisons.

Test Suite		Full NCANDA Battery			WebCNP Tests Only		
Statistic \ Models	8-Factor	Vulnerabilities	Gur+	8-Factor	Vulnerabilities	Gur	Empirical
Model Number	1	2	3	4	5	6	7
N	806	806	806	827	827	827	827
Likelihood ratio tests							
LR χ^2(saturated)	245.87	270.47	268.44	69.65	91.16	81.88	81.97
df (saturated)	79	77	84	27	25	29	30
P-value (saturated)	0	0	0	0	0	0	0
LR χ^2 (baseline)	2372.1	2372.19	2372.16	829.65	829.69	830.1	831.87
df (baseline)	120	120	120	45	45	45	45
P-value (baseline)	0	0	0	0	0	0	0
Information criteria (lower is better)							
AIC	27642.5	27671.9	27664.1	18220.0	18246.9	18231.0	18219.1
BIC	27969.4	28007.7	27968.6	18391.1	18426.9	18393.0	18376.6
Population error (lower is better)							
RMSEA	0.057	0.062	0.058	0.049	0.063	0.052	0.051
90% CI L.L.	0.049	0.054	0.05	0.035	0.049	0.039	0.038
90% CI U.L.	0.065	0.07	0.066	0.063	0.077	0.066	0.064
P(RMSEA<0.05)	0.082	0.01	0.054	0.537	0.073	0.376	0.435
Baseline comparison (higher is better)							
CFI	0.925	0.914	0.918	0.946	0.916	0.933	0.934
Residual magnitude (lower is better)							
SRMR	0.038	0.042	0.04	0.033	0.04	0.037	0.037

especially for continuous performance ($\boldsymbol{\lambda} = 0.34$). The only loading that was weaker was for the emotion recognition on Social Cognition ($\boldsymbol{\lambda} = 0.24$). There was substantial correlation among the latent variables, exception of postural stability (Balance) and the association between Motor Speed and Episodic Memory (Figure 3.1). The highest correlations were observed among Executive Functioning, Complex Cognition and Social Cognition (r's > 0.88; Figure 3.1). There was one Heywood case among the correlations; the point estimate for the correlation between Social Cognition and Executive Functioning was just above one, but had a standard error much larger than the increment that exceeded one (Figure 3.1).



Factor Correlations

	Attention	WorkMem	EpiMem	Abstractn
WorkMem	**0.25 (0.04)**			
EpiMem	**0.22 (0.05)**	**0.42 (0.05)**		
Abstractn	**0.38 (0.04)**	**0.43 (0.04)**	**0.66 (0.05)**	
SocCog	**0.47 (0.12)**	**0.42 (0.12)**	**0.88 (0.18)**	**1.04 (0.20)**

Figure 3.2. Confirmatory factor analysis estimates of the 8-Factor Model for the WebCNP suite of tests.

Model fit improved when the models were fit only to the WebCNP tests (Table 3.2). Although the 8-Factor Model evidenced the lowest BIC among the conceptually motivated models (Models 1-6, Table 3.2), the Gur+ Model's BIC was of similar magnitude (1.9 points higher). The 8-Factor Model demonstrated the best CFI, RMSEA, and SRMR for all the models estimated in this study (Table 3.2). At 0.95 and 0.05, respectively, the CFI and RMSEA values for the 8-Factor reached the borderline of good fit for factor models. Although the model fit statistics of the empirically derived factor model were slightly inferior to the 8-Factor Model, it had a superior BIC (14.5 points lower) and an equivalent AIC (0.9 points lower).

The 8-Factor Model had similar loading strengths as that observed in the Gur+ Model, for freely estimated loadings (Figure 3.2). The single indicator loadings were constrained by the estimated errors. The loading for the continuous exclusion test remained moderate in the 8-Factor model with a $\boldsymbol{\lambda}$ of 0.39. Correlations among the latent variables in this model were generally lower than observed in the Gur+ Model, with the notable exception of the correlations with Social Cognition (Figure 3.2). The correlation for Social Cognition and Abstraction was structured similarly to the Heywood cased observed in the Gur+ model (Figure 3.2).

In summary, when the CFA models were applied to the full NCANDA neuropsychological battery the Gur+ Model was selected based on the objective selection criteria for the full suite; however, it failed to demonstrate strong model fit, as did its competitors. Model fit for all the models applied to only the WebCNP suite of tests improved, however their constellation of fit indices indicated they were at best at the borderline of being well fit. Overall, the best model fit indices were observed for the 8-Factor Model when it was applied only to the conventional WebCNP tests. It also had favorable information criteria statistics. Only the

Empirical Model had a better BIC. The Empirical Model's fit statistics were equivalent to that observed for the 8-Factor Model.

Invariance

Inspection of the salient loading patterns failed to identify substantial divergences in sex-, age-, or time point-specific models. Having evaluated configural invariance, metric and strong invariance by age, sex, and time point were evaluated in multiple-group CFA models. In the multi-group models evaluating age and time point, adding invariance constraints (metric and scalar) resulted in worsening of the models, based on likelihood ratio tests of nested models (*p*'s ≤ .05). Decrements in all the RMSEA values for the metric and scalar models were small (ΔRMSEA < 0.01), however the values all exceeded 0.05 (Table 3.3). This was most pronounced for the age models, which had point estimates near 0.10 (Table 3.3). In other words, RMSEA changes did not indicate notable worsening model fit with the invariance constraints, however the multi-group model fit was at best borderline based on this statistic. CFI patterns for age and time point were similar; for these groupings the metric model did not demonstrate a substantial worsening of the CFIs (ΔCFI ≤ 0.007). However, the scalar models did results in notable CFI decrements (ΔCFI ≥ 0.02). In other words, constraining intercepts to equality in the year and age group models resulted in worsening model fit as measured by CFI. In a partial scalar invariance model (Model A) the four most variant intercepts were freed. These four were for continuous performance, vocabulary, emotion recognition, and grooved pegboard tests. Intercepts for continuous performance were highest at baseline ($\nu = 0.26$) and remained steady at the follow-ups ($\nu \approx 0.16$). Vocabulary and grooved pegboard both decline over time. The intercept for grooved pegboard consistently rises from 0.40 to -0.20 by year three. The joint parameter tests also identified non-invariance by year and age (Table 3.3). The freed parameters did improve the models (ΔBIC’s > 658), but the CFI decrement for time point only reached 0.012. As with Model A, the partial scalar invariance model for age (Model B) insufficiently improved ΔCFI to meet invariance criteria (Table 3.3). The partial invariance model for age also freed the most pronounced variant intercepts, which were for vocabulary, WRAT4: reading, and the n-back task. The intercepts for vocabulary and reading were lower by 0.39 and 0.33 for participants 17 or older as compared to younger participants, respectively. N-back scores were higher for younger participants by 0.53.

Although the model fit differences for the metric invariance models indicated adequate fit, the strongest score test for metric invariance was notable. The score test for the time point metric model was moderately significant ($\chi^2_{(27)} = 50.1$, $p = 0.011$). Inspection of MIs and loadings both identified the most substantial loading variance was for working memory and attention indicators. In summary, pattern (metric) invariance was adequate for the Gur+ Model. Strong (scalar) invariance was not established for the Gur+ Model, but this noninvariance could partially mitigated through the freeing of limited sets of intercept parameters.

The invariance patterns were similar for the 8-Factor model applied to the WebCNP data (Table 3.4). An additional scalar non-invariance pattern was indicated by a marginal ?BIC between the sex-grouped metric and scalar models (Table 3.4). However, none of the score tests indicated that the invariance patterns for sex and age could be distinguished from sampling error (*p*’s > 0.37). The freeing of the intercepts for the measures that demonstrated the greatest non-invariance in the partial invariance models rectified the ?BIC’s in the scalar models grouped by time and age (Table 3.4). This relaxation allowed the intercepts for continuous exclusion to

Table 3.3. Invariance tests for the Gur+ Model in the NCANDA Suite.

Model	χ^2 (df)	P-value	RMSEA (90%CI)	BIC	CFI	SRMR	ΔBIC	ΔCFI	ΔSRMR	Wald P-value	Score P-value
Sex (girls, boys)											
Configural	457.39(175)	< 0.001	0.070(0.063, 0.078)	28294.3	0.876	0.055				0.372	
Metric	471.67(184)	< 0.001	0.069(0.062, 0.077)	28247.4	0.873	0.057	46.9	0.003	-0.002		0.157
Scalar	495.73(193)	< 0.001	0.070(0.062, 0.077)	28201.1	0.867	0.063	46.3	0.006	-0.006		0.588
Year (baseline, year 1, year 2, year 3)											
Configural	1142.29(343)	< 0.001	0.065(0.060, 0.069)	92792.1	0.906	0.049				0.008	
Metric	1197.16(370)	< 0.001	0.063(0.059, 0.067)	92635.8	0.902	0.054	156.3	0.004	-0.005		0.008
Scalar	1470.72(397)	< 0.001	0.070(0.066, 0.073)	92676.7	0.873	0.062	-40.9	0.029	-0.008		0.000
Partial Scalar A	1691.96(456)	< 0.001	0.070(0.066, 0.073)	95486.7	0.9.00	0.079	-2851	0.002	-0.025		0.000
Age (<16.5, >=16.5)											
Configural	638.43(175)	< 0.001	0.090(0.083, 0.098)	28161	0.76	0.131				0.024	
Metric	662.16(184)	< 0.001	0.090(0.082, 0.097)	28130.3	0.753	0.135	30.7	0.007	-0.004		0.021
Scalar	739.99(193)	< 0.001	0.093(0.086, 0.101)	28137	0.717	0.162	-6.7	0.036	-0.027		0.000
Partial Scalar B	667.8(190)	< 0.001	0.088(0.081, 0.095)	28077.6	0.753	0.143	52.7	0.000	-0.008		0.028

Note. Partial Scalar Model A = scalar + Continuous Performance, Emotion Recognition, Word Memory, and Grooved Pegboard intercepts were freed, Partial Scalar Model B = scalar + N-Back, WRAT4: Reading, and Vocabulary intercepts were freed for baseline.

gradually rise 0.36 from baseline through to the 3rd follow-up. Freeing of the intercepts for the continuous exclusion test in addition to facial memory and emotion recognition limited the remaining scalar non-invariance to acceptable levels (Table 3.4).

DISCUSSION

We found support for the Gur+ Model when applied to the neuropsychological accuracy scores from NCANDA. The Gur+ Model was the superior model for the full battery, based on the penalized-likelihood criteria. However, this model's fit should be considered modest because the fit statistics were at best at the margins of acceptable levels (Brown, 2015; Browne & Cudeck, 1993; Hu & Bentler, 1999b). When the suite of test scores entered into the factor model was restricted to the subset of tests obtained from NCANDA's WebCNP tests, the model fit statistics slightly improved. This was expected, in part, because the Gur Model was conceptually developed specifically for the WebCNP (Moore et al., 2015). The results from this study provide some support for the use of factor model configurations consistent with the conceptually motivated Gur Model.

The 8-Factor Model's fit was similar to the Gur+ Model. The difference among these two models is limited to their treatment of executive functioning (Table 3.1); these models had the same factor structure except the 8-Factor Model separated out attention and working memory constructs as factors rather than grouping them into an executive functioning factor. In addition, the conditional exclusion test is placed within the Complex Cognition factor. Differences in models' statistics reflect the relative merit of these configural distinctions.

Whereas the primary model selection statistic (BIC) identified the Gur+ Model for the full suite of NCANDA tests, its superiority over the 8-Factor Model was equivocal based on the

Table 3.4. Invariance tests for the 8-Factor Model applied to the WebCNP test suite.

Model	χ^2 (df)	P-value	RMSEA (90%CI)	BIC	CFI	SRMR	ΔBIC	ΔCFI	ΔSRMR	Wald P-value	Score P-value
Sex (boys, girls)											
Configural	150.26(59)	< 0.001	0.068(0.055, 0.082)	18563.1	0.885	0.042				0.417	
Metric	157.31(64)	< 0.001	0.066(0.053, 0.079)	18534	0.883	0.045	29.1	0.002	-0.003		0.327
Scalar	174.6(69)	< 0.001	0.068(0.055, 0.08)	18526.8	0.867	0.046	7.2	0.016	-0.001		0.566
Partial Scalar C	173.31(68)	< 0.001	0.068(0.056, 0.081)	18530.1	0.867	0.046	3.9	0.016	-0.001		0.454
Year (baseline, year 1, year 2, year 3)											
Configural	295.99(113)	< 0.001	0.053(0.046, 0.061)	61218.3	0.945	0.033				0.072	
Metric	319.84(128)	< 0.001	0.051(0.044, 0.058)	61124.8	0.942	0.04	93.5	0.003	-0.007		0.105
Scalar	640.88(143)	< 0.001	0.078(0.072, 0.084)	61331.2	0.85	0.053	-206	0.092	-0.013		0.010
Partial Scalar D	348.77(134)	< 0.001	0.053(0.046, 0.06)	61107.4	0.935	0.041	17.4	0.007	-0.001		0.110
Age (<16.5,>=16.5)											
Configural	217.85(59)	< 0.001	0.09(0.077, 0.103)	18476.0	0.776	0.068				0.401	
Metric	222.85(64)	< 0.001	0.086(0.074, 0.099)	18444.0	0.775	0.071	32	0.001	-0.003		0.414
Scalar	239.71(69)	< 0.001	0.086(0.074, 0.098)	18424.0	0.759	0.072	20	0.016	-0.001		0.376
Partial Scalar E	227.9(67)	< 0.001	0.085(0.073, 0.097)	18436.5	0.773	0.074	7.5	0.002	-0.003		0.438

Note. Partial Scalar Model C = scalar + Matrix Reasoning intercepts were freed, Model D = scalar + Emotion Recognition, Facial Memory, and Continuous Exclusion intercepts were freed, Model E = scalar + Continuous Exclusion and Visual Object Learning intercepts were freed.

small difference in models' BIC values. However, the 8-Factor Model's fit statistics were slightly superior when applied to the WebCNP suite of tests. Based on the statistical evaluation conducted here, researchers could justifiably interchange these models in order to optimally align the latent factors with the substantive targets of their investigations. This is a reasonable conclusion when the alternative models are equally supported in the descriptive literature (i.e., *phenomenological* models), and there is not a well-formed accepted theory to derive a focusing model (Boniolo, 2007; Cummins, Pitpitan, Reed, & Zuniga, Under Review; Chapter 2). Where alternative models are available, selection should optimize the trade-offs between realism, precision, and generality for the particular scientific question under investigation (Matthewson, 2011; Levins, 1966); under the perspective that no factor model is correct, the selection of an optimal approximating model should be based on context specific scientific considerations (Preacher & Merkle, 2012).

The 8-Factor Model has several advantages over the Gur Model for application in the NCANDA study. The first is that it separates out neuropsychological constructs (working memory and attention) that have each been identified as being vulnerable to alcohol's biological effects during adolescence (Jacobus & Tapert, 2013; Spear, 2018). Second, this configuration is more aligned with the factor structure empirically identified for WebCNP efficiency scores (Moore et al., 2015; James et al., 2016). This is because the 8-Factor model places the continuous exclusion test with verbal reasoning and matrix reasoning, which tap the Complex Cognition factor of the Gur Model. Finally, in some contexts the lower correlations among the latent constructs in the 8-Factor Model will provide improved discriminant capacity (Farrell, 2010). Although the findings presented in this study indicate that either of these two models can be reasonably be justified for use as a measurement model based on its individual model fit and conceptual underpinnings, the 8-Factor Model may prove to be of greatest utility for addressing the aims of the NCANDA study.

An empirically derived factor model was also evaluated in the current study. Moore et al. (2015) reported an exploratory factor analysis model based on accuracy scores that we evaluated in a CFA framework. This model collapses the highly correlated Executive Functioning and Complex Cognition factors found in the Gur Model. It also parses the episodic memory and social cognition items differently (Table 3.1, Empirical Model). Providing some corroboration for Moore et al.'s earlier finding, the Empirical Model was the best model for the NCANDA WebCNP suite based on BIC and was equivalent to the 8-Factor Model based on AIC (Table 3.2). The metric version of this model never converged in any of the bootstrap iterations, so only the configural structure is supported by the current study’s findings. NCANDA's WebCNP battery was modified from the standard version, in that it included two supplemental delayed recall tests and a modified nonverbal reasoning test and dropped a social cognition test. These differences could partially explain the discrepancies between the original loadings and loadings estimated in the current study that resulted in the lack of convergence.

Moore et al. (2015) and James et al. (2016) provided the only other published factor analysis of the standard WebCNP battery. The findings presented here are consistent with the earlier reports. Moore estimated a CFI of 0.95 for the Gur Model applied to efficiency scores. This was higher (+0.032) than what was observed in this study. Support for the Gur model's application is further strengthened by a number of favorable invariance findings. Based on CFI differentials, the Gur+ Model did not evidence substantial violations of metric invariance assumptions (Sellbom & Tellegen, 2019). However, there was evidence of at least mild intercept variance for time point and age, which could not be entirely mitigated in the partial scalar models. The scalar non-invariance indicates that some of the neuropsychology scores varied between groups in a way that diverged from the patterns of variability in the mean of the latent variable (Marsh et al., 2018).

Two of the neuropsychology tests contributing to scalar invariance violations, were also associated with a moderately significant joint (score) test indicating some metric invariance violations across time points. These were the continuous performance test and the short fractal n-back test, which were designed as assessments of attention and working memory, respectively. These patterns may be partially attributable to differential practice effects that disrupt the correlation structure of the measurement battery. A prior report based on the NCANDA sample found these tests to be most sensitive to practice effects (Sullivan et al. 2017). Rather than using raw test scores, application of test specific scores that are adjusted for test-specific practice effects could alleviate some temporal non-invariance. Additionally, neurocognitive domains (e.g. executive functioning and memory) and their component processes develop at different rates during childhood and adolescence, which could result in age related non-invariance (Thompson et al., 2019; Poon, 2017). However, differential development of neuropsychological functioning would not be expected to contribute to non-invariance where the structure within factors is stationary as people age. Gur et al. did identify developmental patterns that were specific to individual WebCNP tests (2012). For example, they found substantial and distinct improvements between childhood to early adulthood on the continuous performance test. This could necessitate the use of developmentally normed scores for longitudinal use of measurement models applied to WebCNP scores, if the effects are found to be substantial in particular research settings. Sensitivity analyses could be used to assess the importance of the observed magnitudes of scalar invariance and the benefit of applying mitigations such as multiple group factor analysis alignment (Asparouhov & Muthén, 2014; Marsh et al., 2018). More importantly, distinct divergence in the developmental trajectories of individual indicators for putative factors should call into question the meaning of latent constructs created by factor analysis and our conceptualization of their connections to changes in the brain. It is up for debate, but some alternative approaches, such as network models, appear, at least on the surface, better able to address these complexities (van der Maas, Kan, Marsman, & Stevenson, 2017).

One of the current study's limitations is the absence of an evaluation of potential changes in neuropsychological architecture as a result of alcohol exposure that would be represented as metric non-invariance. This is an important consideration in the context of the NCANDA project's focus on effects of alcohol exposure, where such non-invariance might not be considered a statistical nuisance but instead be of material interest. The study design favored substance use naïve participants at baseline (Brown et al., 2015), with few alcohol exposed participants recruited and only a small number transitioning into heavy episodic binge drinking patterns over the period analyzed in this report (Brumback, Thompson, Cummins, Brown, & Tapert, 2020). The current study is not yet powered to detect subtle non-invariance associated with drinking exposure, but should be of substantial interest to future NCANDA investigators as the sample matures.

An important consideration of the current work is the limited number of indicators for some latent factors. Although, using single indicators can create psychometric (Schmitt et al., 2018) and computational challenges (Bollen, 2014), their use may be necessitated by the underlying neuropsychological architecture and its match with the available measurements. Indeed, the NCANDA neuropsychological battery was designed to efficiently survey broad domains of neuropsychological functioning. This left some domains covered by a limited number of tests. In this study, when factors with single indicators were included, we used independent estimates of the error variance to determine the factor models in this study (Brown, 2015). Irrespective of the quality of these variance estimates, this mitigation does not address the potential for construct underrepresentation (Kaplan & Saccuzzo, 2008), which cannot be mitigated without a study design change.

The study provides several contributions. It provides an independent corroboration of the Gur model and supports the adequacy of the Empirical Model to capture the factor structure of a computerized neuropsychological test battery that is easy to deploy and administer over the internet. Further, this is the first publication that includes formal invariance testing of the WebCNP. The study also advances the approach of evaluating confirmatory factor models in a comparative framework, rather than looking at individual models in isolation. The comparisons in this study provide evidence that the conceptually derived model targeting neuropsychological processes posited to be most vulnerable to alcohol's effects in adolescents (Vulnerabilities Model) was the most inferior model evaluated. As noted above, the 8-Factor model was found to be sufficient and evidenced slight superiority in model fit to the alternative models when applied to the WebCNP Suite of neuropsychological tests. Although, there were marginal Heywood cases observed in the leading models, which is possible even if models are correctly specified when estimation is based on maximum likelihood (Van Driel, 1978; Ximénez, 2006), a confidence interval for these cases broadly covered admissible ranges. Thus, use of either of the leading models should be conducted with the recognition that Social Cognition is highly correlated with complex cognition and re-specification of the model to accommodate this aspect of the relationship could potentially improve some characteristics of the models.

In the absence of an integrated theory of brain and neurocognitive functioning that functions as a well-formed engrained theory we will continue to work with factor models that are influenced by research inertia and indeterminant competition among alternatives models (see Cummins et al. 2020 for a discussion of engrained Theory). Work by authors such Patt et al. (2017) on the structure of neuropsychology and others who are investigating the connections between alterations in brain and neuropsychological functioning as a consequences of exposure to substance use, infectious disease, and trauma (Squeglia, Spadoni, Infante, Myers, & Tapert, 2009b; Bava, Jacobus, Mahmood, Yang, & Tapert, 2010; Squeglia et al., 2012a; Mota et al., 2013; Worbe et al., 2014; Banca et al., 2016; Yadav et al., 2017) provides the empirical foundations necessary to achieve such theory. Until a theory is fully developed, some exploratory efforts to aptly describe the variance patterns within neuropsychological batteries should continue, even where a confirmatory factor model has been previously developed and found to fit in independent samples. These efforts can operate in a Bayesian framework (e.g., (Thompson et al., 2019) or in an exploratory framework where series of exploratory factor models are compared (e.g., (Patt et al., 2017). The recommendations to continuously interweave exploratory factor analysis and confirmatory factor analysis into model building (see (Schmitt et al., 2018) would be most appropriate when conducting phenomenological modeling, including in research contexts like the NCANDA project where the statistical superiority of alternative models is equivocal.

ACKNOWLEDGEMENTS

The data presented here are based on the National Consortium on Alcohol & NeuroDevelopment in Adolescence (NCANDA) Data Release NCANDA_RELEASE_3Y_REDCAP_MEASUREMENTS_V02 which was made possible through the NIH grants AA021695, AA021692, AA021696, AA021697, AA021690, AA021681, and AA021691. Dr. Moore is supported by NIMH grant MH117014 and the Lifespan Brain Institute (LiBI) of the Children's Hospital of Philadelphia. NCANDA data can be obtained through the National Institute of Alcohol Abuse and Alcoholism (NIAAA) at https://www.niaaa.nih.gov. The variable indexing families will be included in future releases. The authors wish to recognize Ms. Sonja Eberson for their helpful comments on early drafts of this manuscript.

Chapter 3 has been submitted for publication and is under review: Cummins, K. M., Pitpitan, E. V., Brown, S. A., Clark, D. B., Brumback, T., Trim, R. S., & Tapert, S. F. Comparison of factor analysis models applied to the NCANDA neuropsychological test battery. *Journal of Clinical and Experimental Neuropsychology*. Kevin Cummins was the primary investigator and author of this paper.

CHAPTER 4: MEDIATION OF HEAVY EPISODIC DRINKING EFFECTS ON ACADEMIC PERFORMANCE THROUGH NEUROCOGNITIVE FUNCTIONING

ABSTRACT

The role of neurocognitive decrements attributable to heavy episodic drinking in the association between adolescent alcohol use and academic functioning is not well characterized. This study leverages 419 participants from the NCANDA study, which is a longitudinal study of youth neurodevelopment. Participants were measured on their school grades, substance use exposures, and neurocognitive functioning up to two times a year for four years. Linear mixed models with propensity score weighting were employed to evaluate the mediational role of neurocognitive functioning. Mild associations between heavy episodic drinking and grade point average were observed. The strongest associations were for executive functioning and working memory. Decrements attributable to heavy episodic drinking were not observed. Mediation could not be established. The low level of exposures experienced reported by the sample limited the study's capacity to detect effects. The analytic framework developed and employed in this study provides service to future studies of the pathway under investigation.

INTRODUCTION

Adolescence is an important period of neuromaturation and a sensitive period for persistent alterations to the brain's reward circuitry in response to alcohol exposure in animal models (Crews et al., 2016; Salling et al., 2018; Spear, 2018); drinking onset during this period is well-established risk factor for alcohol use disorder in adult humans (Grant et al., 2006; Grant & Dawson, 1997). Emerging evidence suggests that normative patterns of alcohol exposure experienced by community youth can result in alterations in neuroanatomical development (Pfefferbaum et al., 2017; Squeglia et al., 2009b). Neurocognitive deficits associated with adolescent alcohol exposure include verbal learning (Mahmood et al., 2010; Sneider et al., 2013; Nguyen-Louie et al., 2016), visuospatial processing (Tapert & Brown, 1999), executive functioning (Mota et al., 2013), attention (Tapert & Brown, 1999; Tapert et al., 2002) and memory (Mota et al., 2013; Hanson et al., 2011; Tapert et al., 2002; Nguyen-Louie et al., 2016). The implications for psycho-social functioning of alcohol-induced neuromodulation are just now being confirmed in large-scale longitudinal community samples (Brown et al., 2015; Jernigan et al., 2018; Akshoomoff et al., 2014). These studies are designed to provide a more definitive evaluation of the consequences of alcohol exposure during this period of potential heightened sensitivity.

Even if the consequences of alcohol use during adolescence are subtle at the individual level, implications could be substantial at a societal scale. Alcohol remains the most common substance used by adolescents in the United States. The most prevalent pattern of alcohol consumption among adolescents is heavy episodic drinking (HED), which is commonly referred to as “binge” drinking (Chung & Jackson, 2019). Fourteen percent of American high school seniors reported at least one HED event in the prior two weeks in the 2019 Monitoring the Future survey (Johnston et al., 2019). Even though reports of lifetime drinking have declined by 15 percentage points among high school seniors over the last 30 years, the majority leave high school with alcohol experience (Twenge & Park, 2019).

Alcohol use among youth remains a global public health concern. Globally, alcohol is the strongest risk factor for a worsening of disability adjusted life-years among 15-24 year-olds (Gore et al., 2011). Alcohol is an attributable factor in 17% of deaths among 15-19 year-olds in North America. The rate is even higher in Europe (20%). In most developed western nations adolescent alcohol involvement (i.e., lifetime frequency, lifetime and recent HED) exceeds that experienced by adolescents in the United States (Ahlström & Österberg, 2004). Additionally, culminating evidence implicates adolescence as a sensitive period for the potentiating effects of alcohol exposure to later risk of addiction, a major contributor to the global burden of disease (Rehm et al., 2009; Rehm & Imtiaz, 2016; Rehm et al., 2017; Whiteford et al., 2013). Youth alcohol consumption has a well-established association with depression and suicidal behaviors (Galaif, Sussman, Newcomb, & Locke, 2007; Windle et al., 2008). Relational problems, risky sexual behavior, and doing things that were later regretted are also substantially elevated among adolescents who use alcohol (Windle et al., 2008). Adolescent alcohol use is also predictive of poorer adult psycho-social outcomes, including substance use disorders, antisocial behaviors, and depression (Brown et al., 2008). The mechanisms between these associations appear complex and remain to be fully elucidated.

Alcohol consumption also has a well-established negative association with academic functioning, performance, and attainment among high school students (Cox et al., 2007; Lopez-Frias & Fernandez…, 2001; Singleton & Wolfson, 2009; Friedman & Glickman…, 1985; Latvala et al., 2014; McCluskey et al., 2002; Bergen et al., 2005; Engberg & Morral, 2006; Balsa et al., 2011; Porter & Pryor, 2007; Bachman et al., 2008). High school dropouts are three times more likely to have a history of binge drinking in early adolescence when compared to youth who attain more than 3 years of college.(Bachman et al., 2008) HED among adolescents is prospectively correlated with lower grade point average ($r > -0.15$)(Bachman et al., 2008). Although common pre-existing confounders (e.g. familial environment) may explain part of the association, the association is thought to include causal mechanisms (Bachman et al., 2008). Some have postulated that adolescent substance use mutually facilitates the development of social networks with high levels of antisocial attitudes, with both determinants interacting to contribute to reductions in academic engagement and performance via shared and independent causal pathways (Bachman et al., 2008). The proximal mediator in this pathway is school disengagement, a dimension of academic functioning. This is just one of many putative co-occurring mechanisms that are suggested as explanations for lower academic success subsequent to alcohol engagement during this developmental phase (Bachman et al., 2008). Conversely, low scholastic success can precipitate the escalation of alcohol use (Crum et al., 2006; Schulenberg et al., 1994; Bryant et al., 2003). Regarding the causal mechanisms, Bachman et al. (Bachman et al., 2008) conclude that: 1) prior common causes are operating, 2) academic success modulates the risk of substance use, and 3) substance use impacts academic success. They postulate that the effect of substance use exposure, including alcohol, is the weakest causal link. However, this pathway has not been adequately quantified.

The literature has not disentangled, nor resolved, the various mechanisms behind the effects of alcohol on academic functioning. A substantial body of work has accumulated addressing both the association of neurocognitive functioning and academic success and alcohol use and academic functioning. Additionally, a coalescing body of work is developing on the association between normative youth alcohol exposure and neurocognitive functioning. Although this work has progressed over the last decade, these three areas have not been sufficiently tied together despite calls from King, Meehan, Trim & Chassin (2006) to establish the relative importance of neurocognitive impairment attributable to alcohol exposure as a mechanism leading to poorer academic outcomes.

Prospective studies have found correlations ranging from 0.40 to 0.63 between baseline neurocognitive assessment scores and measures of educational attainment (Jencks, 1979). Grade point average and measures of neurocognitive functioning are similarly correlated (r’s = 0.4-

0.7)(Mackintosh & Mackintosh, 2011). When cognitive ability was measured broadly with relative high reliability, the estimated prospective correlation with standardized academic scores has been found to reach 0.81 (Deary et al., 2007). Although concerns that many early cognitive tests strongly reflected academic achievement, there are measures that are relatively delineated from scholastic skills and experiences (Kamphaus, 2005). Further, longitudinal research suggests that cognitive functioning more strongly influences later academic success, than academic performance influences subsequent measures of cognitive functioning (Watkins et al., 2007).

The aim of this paper is to assess the causal pathway that runs from alcohol use through neurocognitive functioning to academic performance. Neurocognitive domains investigated in this study were limited to executive functioning, memory, and attention which are functions thought to be among the most sensitive to alcohol during adolescence (Lees, Meredith, Kirkland, Bryant, & Squeglia, 2020; Spear, 2018). The associations between measures of neurocognitive performance and prospective grades were first evaluated. Then the causal effects of HED on each of the neurocognitive domains used in this study were estimated. Finally, the incremental effect on neurocognitive functioning that is attributable to HED was used to predict longitudinal changes in student grades.

MATERIALS AND METHODS

Sample

Data originate from the NCANDA project. NCANDA is a longitudinal study of youth that employs a modified accelerated longitudinal design (Miyazaki & Raudenbush, 2000; Galbraith et al., 2017; Brown et al., 2015). Recruitment and enrollment were structured to result in samples demographically representative of each of the five metropolitan catchment regions, with enrichment of participants at risk for lifetime substance use disorder; 23% of this study's recruits were considered high risk for a lifetime diagnosis of alcohol use disorder based on a family history of alcohol use disorder, externalizing or internalizing symptoms, or consumption of alcohol before age 15. Exclusionary criteria included magnetic resonance imaging contraindications (e.g., permanent metal in the head such as dental braces), neurotropic medications, serious medical problems, major Axis I disorders, pervasive development disorder, uncorrectable vision or hearing impairment, and lack of English fluency. As details of the study design were previously published (Brown et al., 2015), a sketch of the protocol is presented here. The attributable effects of HED on neurocognition are expected to be subtle based on prior findings so the identification of mediation effects was substantially advantaged by the use this study's large nationally representative sample.

The baseline observation for this study is the first-time participants reported enrolment in high school. They were followed until their first post-high school assessment, or until the participant was older than 18 years. Participants provided interviewer administered self-report assessments approximately every six-months. Coinciding with an interview, neurocognitive functioning as assessed once a year. The protocol was reviewed and approved by the human research protection program at each participating university.

Measures

Neuropsychological functioning was assessed with the Web-based Computerized Neuropsychological Battery (WebCNP) (Gur et al., 2010). WebCNP assesses a range of neurocognitive domains and provides scores that integrate performance accuracy and speed, where the speed–accuracy trade-off is represented in efficiency scores (Gur et al., 2010). The WebCNP battery was slightly modified from the standard WebCNP distribution to optimize the sensitivity of the battery to detect the effects of alcohol exposure. For example, to better distinguish effects on memory function, a delayed-recall (Delis et al., 2003) version of the delayed Visual Object Learning, Face Memory, and Word Memory tests were included in the NCANDA study (see Sullivan et al., 2015). Efficiency scores were employed in all analyses except for the Continuous Performance Task, where accuracy scores are employed. Efficiency scores were not standard for this task (Neuropsychology Section at University of Pennsylvania, 2012). The three domains of functioning explored in this study are memory, attention, and executive functioning.

Episodic memory is assessed with Visual Object Learning, Face Memory, and Word Memory WebCNP measures. In these assessments, participants are shown 20 stimuli (faces, words, or three-dimensional Euclidean shapes) then approximately 25 minutes later are shown a series of stimuli where the original images are intermixed with 20 distractor images. Participants respond to each stimulus by indicating if the level of their belief that they had been shown the image during the target presentation in a 4-point response scale: definitely yes, probably yes, probably no, and definitely no.

Attention is assessed with the Continuous Performance Task. In this task the participant must press the spacebar whenever a collection of lines within a stimulus form a complete number in one type of trial or a complete letter in the alternative trial type. Each trial lasts for 1.5 min. Each stimulus is displayed for 300 ms and is followed by a blank screen for 700 ms, giving the participant 1 sec to respond to each trial.

Executive Functioning is assessed with a factor score from a factor model that loads the Continuous Performance Task and the Short Fractal N-back task. The Short fractal N-back (N-back) test measures working memory. Participants view fractal designs displayed on the computer screen and indicate the “target design.” There are three trial types. During the 0-back, the target design is designated before the trial and the participant responds each time they see it. For the 1-back and 2-back the target design is indicated by the repetition of a design, with the participants responding when they see a design for the first time for 1-back or the second time for 2-back. In all trials, the participant has 2,500 ms to respond.

Extensive methodological and statistical descriptions of the neurocognitive assessment protocol and scores can be found in Sullivan et al. (2016). Distributions of neurocognitive scores were roughly normally distributed, thus application of general linear models were deemed appropriate without the need for transformation of variables. Gur et al. (2010) established construct validity and reliability of the WebCNP assessments with a community sample (n ~ 10,000) that included a wide age range (8 to 90 year-old).

Demographics, including sex and age, parental socio-economic status (income, occupation and educational attainment), pubertal development status, were obtained during the baseline interview via self-report as described in Brown et al. (2015). Familial relationships among participants was also identified through proband and parental self-report assessments.

Academic behaviors, attitudes, and performance were assessed via proband self-reports. School connectedness was assessed with a single item from the California Healthy Kids Survey (Austin, Polik, Hanson, & Zheng, 2018), which has good agreement ($\alpha = 0.86$) with other school connectedness items in the CHKS. The item prompt is, “I feel like I am part of this school.” It has a five-point response scale: strongly disagree, disagree, neither agree nor disagree, agree, strongly agree. Peer group attitudes toward school and school engagement was assessed with the Peer Group Deviance measure that was adapted from Monitoring the Future (Bachman, Johnston, O’Malley, Schulenberg, & Miech, 2015). It asks, “Friends are people who you see regularly and spend time with in school and outside of school. How many of your friends: Skip or cut school a lot?" This item's response scale is none, a few, some, most and all, which are coded as 0, 1, 2, 3, and 4, respectively. Academic performance was measured with grade point average. This was obtained during a self-report interview with an item that asks, "What are your most recent grades like (estimate GPA on a 4-point scale)?" The proband report of grades had a moderate reliability when compared with the parental report of grades ($\alpha = 0.78$).

Substance use exposure was assessed through the Customary Drinking and Drug Record (CDDR) during in-person interviews (Brown et al., 1998). The CDDR assesses the frequency of heavy episodic drinking episodes (HED) (4/5+ drinks on an occasion for girls/boys) during the prior 30 days and previous 12 months, in addition to other dimensions of alcohol consumption. Use frequency (number of days) of alcohol and cannabis, during the same reference periods are also obtained. In addition, days of nicotine use in the prior 30 days was reported. Participants also reported their maximum number of standard drinks during a single event and the mean number of drinks per event during the prior year.

Statistical Analysis

Summary Statistics

Means, standard deviations, maximum, and minimums of all of the substance use exposure variables, covariates, and neurocognitive variables were computed for the baseline observation and the final observation for each participant. Descriptive statistics were stratified by lifetime HED experience for the final observation. Inferential statistics are not included with the descriptive statistics, which is a practice consistent with the STROBE guidelines (Vandenbroucke et al., 2007).

Confirmatory Factor Analysis

The Executive Functioning factor's configuration was conceptually motivated by Gur and colleagues (2010) treatment of the assessment battery and empirically supported by data presented by both Moore et al. (2015) and Cummins et al. (2020). Model parameters were estimated from the study sample. Factor scores were computed using Bartlett's method (Estabrook & Neale, 2013).

Statistical Models: Association Between Neurocognitive and Academic Performance

The association between neurocognitive functioning and grade point average was evaluated with a Pearson product moment correlation as a supplement to linear mixed effects models. The mixed effects models were included to address the nonindependence of observations through the inclusion of random effects for participants and families. Grade point average was entered as the outcome variable and the neurocognitive scores entered as predictor variables in separate models. Neurocognitive scores were used from the assessment immediately prior to the interview, when the grade point average reports were obtained (~ 6 months after the neurocognitive assessment). Semi-standardized coefficients were reported, where the grade point average was retained on its normal 4-point scale and each of the neurocognition scores were standardized based on the sample standard deviation. These unadjusted models were supplemented with higher order linear mixed effects models and generalized additive models (GAM) to explore the potential for curvilinear associations (Hastie & Tibshirani, 1990; Wood, 2017).

Statistical Models: Association Between Heavy Episodic Drinking and Neurocognitive Performance

The modeling approach applied to the relationship between heavy episodic drinking and neurocognitive performance was based on generalized linear models with random effects and propensity score weighting. Propensity score anchored analyses provide a statistically efficient means to address covariate imbalance, where causal inference is the analytic goal (Rosenbaum, 2009; Gelman & Hill, 2006). The type of propensity scores employed here are Generalized Propensity Scores (GPS) estimated with generalized boosted models (McCaffrey et al., 2013; Zhu, Coffman, & Ghosh, 2015; Schuler, Chu, & Coffman, 2016), which automatically include higher order model specifications (i.e. interactions and curvilinear associations among potential confounders). Generalized boosted models were executed via the twang package in R (Parast et al., 2017). Stopping rules were based on the standardized bias and the Kolmogorov-Smirnov (KS) statistic. Evaluation of the success of confounder balancing through propensity score weighting was addressed using the approaches outlined by Ridgeway (2014) including the evaluation of standardized differences after weighting, Kolmogorov-Smirnov tests for the comparison of the unweighted and weighted distributions, and stratified propensity score visualizations. Propensity score weights were used to calculate partially-double robust average treatment effect for the population (ATE) of heavy episodic drinking on each neurocognitive outcome. The largest RI after weighting was for developmental markers. Because age-based trajectories were the basis of the modeling in this study, some of confounding that is due to developmental differences has been adjusted out of the effects estimates as complementary modeling action to the propensity score weighting. This makes the estimators partially-double robust.

The propensity scores were used as weights in linear mixed effects models. The outcomes in the models were the neurocognitive scores and the predictor variable was the number of heavy episodic drinking events in the prior year. The covariates were second order age effects, days since the last binge episode, and an indicator variable for practice effects. The last binge episode variable was transformed with an inverse hyperbolic sine function that was tuned so that its (recovery) effect is halved by one month and tapered to an asymptote near 10 months. Neurological recovery during abstinence among adolescents has been described to occur on a time scale as short as one month (Brumback et al., 2015b; Winward, Hanson, Bekman, Tapert, & Brown, 2014). Because notable practice effects have been detected after the first exposure to the neurocognitive assessment battery in this NCANDA sample (Sullivan et al., 2017), a variable indicating if the neurocognitive scores were obtained from participants' first exposure to the neurocognitive assessment was included. Interactions with heavy episodic drinking frequency were included with all terms except practice effects. Three nested models were compared with likelihood ratio tests. The base model included all the terms except the drinking exposure variables, the second model added the HED terms, and the final model added the binge recency variable to the second model. The final model was tested against the second model. Effects attributable to HED were computed by calculating the difference in the expected value for each neurocognitive score, based on each participants' trajectory estimated from the models and comparing that to the expected value under a counterfactual drinking status of no HED during the assessment period (see Figure 4.1). The difference between these expected values is the attributable effect of HED on neurocognitive performance. A final set of analyses utilized the attributable effects of HED in linear mixed effects models as predictor variables. Prospective grade point average was the outcome variable. Family and participant were included as random effects. Higher order models and GAMs were included in supplementary exploratory analyses to evaluate the potential for curvilinear associations. All analyses were run in R (version 3.6.2) other than weighted linear mixed models which were run in Stata (version 16.0).

RESULTS

Descriptive Statistics

At the baseline, ages ranged from 12.4 to 17.4, with a mean of 14.9, which was 2.4 years younger than the participants at their final observation. The mean number of observations per participant was 2.8 (SD = 0.76). The sex distribution was evenly split (Table 4.1). Approximately, a quarter of the participants were identified as high risk for substance use disorder, and by the final observation 38% of those who had a lifetime experience with binge drinking were high risk whereas only 19% were high risk among those without an exposure to binge drinking (Table 4.1). Bingers also demonstrated small elevations in the truancy proxy

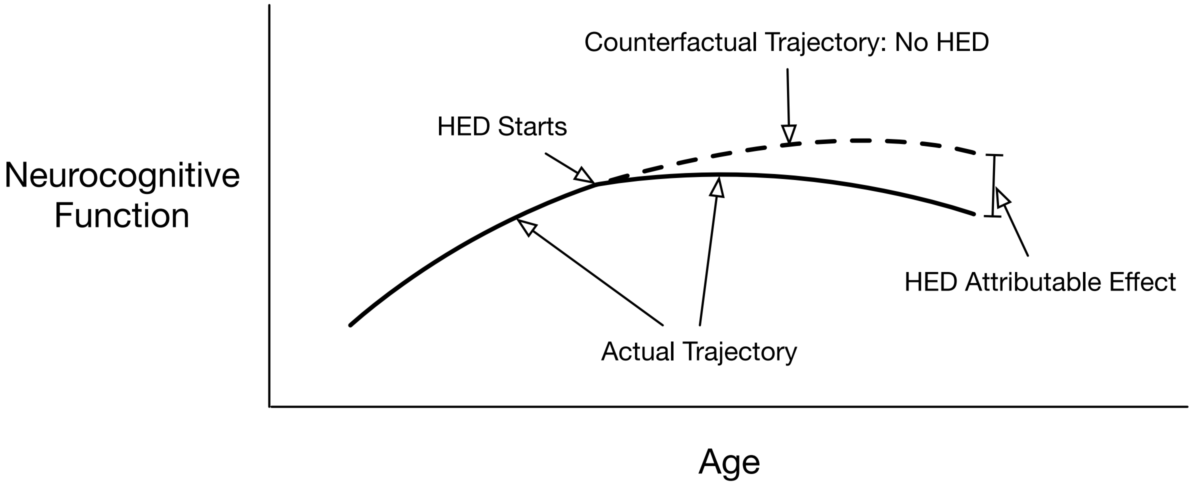


Figure 4.1. Conceptual representation of how the heavy episodic drinking (HED) attributable effects on neurocognition were estimated. Counterfactual trajectories were based on predictions obtained from the estimated models with the HED exposure set to zero for each observation.

variable, but had similar school connectedness scores (Table 4.1). Substance use at baseline was limited, with only 4% reporting a binge episode prior to the baseline observation. Cannabis use among existing bingers was substantially higher among bingers where 78.1% reported using cannabis in the prior year as opposed to only 15.1% of the binge naïve participants. However, 69.3% of the cannabis users reported using cannabis less than one time a month, on average in the prior year.

There were modest correlations among the neurocognitive test scores included in the study. The highest correlation (r = .28, 95% CI: .19, .37) was observed between Word Memory and Face Memory. However, the Executive Functioning factor score was highly correlated with its indicators (r's = .75, .79; 2-back and Continuous Performance) and moderately correlated with the remaining test scores (r's = .24-.33; Table 4.2). Continuous Performance was the least correlated with other test scores with correlations ranging from 0.12, with Visual Learning, to 0.23, with Face Memory.

Propensity Scores

The Kolmogorov-Smirnov scores reached an asymptote by 10,000 iterations. The strongest relative influences (RI) in the GBM were pubertal development scores (RI = 49.2) and socio-economic status (RI = 28.1). The propensity weighting rectified the absolute standard difference of all of the potential confounders that were initially greater than 0.30, resulting in a maximum of 0.27 for pubertal development after weighting. The domain for the propensity scores differed between lifetime bingers and binge naïve participants, such that there was poor overlap among the lowest scores. The region of poor overlap was addressed by dropping sixteen participants with the lowest propensity scores, which were all binge naïve cases. Kolmogorov-Smirnov distributions also demonstrated imbalance rectification and approximated a randomized distribution with the weighted values (data not shown). The effective sample size for participants reporting HED and those who did not are 142.4, and 211.3, respectively.

Table 4.1. Summary statistics.

Variable	Initial Observation (N=416)	Final Observation No HED Exposure (N=320)	HED Exposed (N=96)
Sex (male)	202 (48.6%)	155 (48.4%)	47 (49.0%)
Age (years)	14.9 (0.935)	16.9 (0.785)	17.2 (0.641)
Pubertal development score	3.08 (0.591)	3.46 (0.433)	3.50 (0.398)
Socio-economic status	91.4 (14.4)	89.2 (14.6)	97.4 (12.1)
Race/Ethnicity			
Amer. Indian	1 (0.2%)	1 (0.3%)	0 (0%)
Asian	30 (7.2%)	26 (8.1%)	4 (4.2%)
Pacific Islander	2 (0.5%)	2 (0.6%)	0 (0%)
Black/African-Amer.	49 (11.8%)	48 (15.0%)	1 (1.0%)
White/Caucasian	288 (69.2%)	208 (65.0%)	80 (83.3%)
Other	46 (11.1%)	35 (10.9%)	11 (11.5%)
High risk for substance use disorder	97 (23.3%)	61 (19.1%)	36 (37.5%)
School connectedness	3.97 (0.996)	3.98 (0.957)	4.11 (1.14)
Cutting classes score	0.216 (0.521)	0.374 (0.705)	0.613 (0.738)
Lifetime binging experience	16 (3.8%)	0 (0%)	96 (100%)
Mean number of drinks per event (prior year)	0.296 (1.09)	0.281 (0.682)	4.27 (2.15)
Maximum number of drinks per event (prior year)	0.422 (1.51)	0.344 (0.842)	7.43 (3.85)
Mean days per month drinking (prior year)	0.057 (0.32)	0.0003(.004)	0.043 (0.25)
Days since last binge	97.4 (105)	NA (NA)	69.3 (65.4)
Mean days per month using cannabis (prior year)	0.113 (1.34)	0.208 (1.37)	2.96 (6.10)
Days using nicotine in the prior 30 days	6.60 (13.1)	10.7 (16.8)	6.29 (8.49)

Note. Cells are means (standard deviations) or counts (percentage).

Neurocognitive Functioning and School Grades

Neurocognitive scores were positively associated with grade point average. The evidence was strongest for the association between Executive Functioning and Continuous Performance scores (Table 4.3). The semi-standardized slopes between grade point average and the standardized neurocognitive scores ranged from 0.021 to 0.059. Executive functioning had the strongest association with grade point average (Table 4.2), such that for every standard deviation increase in the factor score grade point average increased by 0.021 (95% CI: -0.002, 0.043). However, this may underrepresent the strength of the association across parts of Executive Functioning's domain. The GAM model fit demonstrates a steeper slope above the mean of Executive Functioning score and a mild attenuation below about one-half standard deviation below the mean (Figure 4.2). Neurocognitive scores explain only a small proportion of the overall variance in grade point average. The raw Pearson product moment correlations ranged

Table 4.3. Model estimates for the association between grade point average and neurocognitive performance.

Measure	r	p-value	Coefficient	SE	df	p-value
Executive Functioning Factor Score	0.24	< .0001	0.059	0.011	1015.3	< .0001
Continuous Performance Test	0.16	< .0001	0.057	0.011	925.6	< .0001
Delayed Face Memory Task	0.11	0.050	0.038	0.012	1091.3	0.0023
Delayed Visual Object Learning Test	0.15	< .0001	0.031	0.012	1080.3	0.0146
Delayed Penn Word Memory Task	0.11	0.040	0.027	0.012	1096.2	0.0376
2-back Task	0.21	< .0001	0.021	0.012	1020.2	0.0713

r is the Pearson product moment correlation between grade point average and each neurocognitive performance score. Coefficient is the semi-standardized coefficient from the linear mixed effects model where only the neurocognitive score was standardized. Models are adjusted for age and practice effects. SE is the standard error of the coefficient.

from 0.10 to 0.24, which translates into a maximum of 5.76% of the variance explained based on a simple linear correlation.

Heavy Episodic Drinking and Neurocognitive Functioning

Evaluation of the effects of HED on trajectories of neurocognitive performance failed to identify consistent patterns when adjusting for age-related changes (Table 4.4; p’s > 0.115). The most discernible pattern among this set of models was for the visual object learning task with a p-value that only reached 0.12. As with all of the models, when adjusting for the days since the last binge event the statistical significance worsened substantially (p = 0.558). With the exception of the Continuous Performance Test, the majority of attributable effects for the neurocognitive scores estimated from these models were near zero (Figure 4.3). The largest decrements at the individual observation level was for the measure of attention (Continuous Performance Test) and visual learning (Visual Object Learning and Memory Test), however 30.0

Table 4.4. Model statistics for the association between heavy episodic drinking predicting to neurocognitive performance.

	Base Model			Recovery Adjusted		
Measure	χ^2	df	p-value	χ^2	df	p-value
Delayed Visual Object Learning Test	4.33	2	0.115	0.59	2	0.745
2-back Task	3.83	2	0.147	1.17	2	0.558
Delayed Face Memory Task	1.35	2	0.510	0.7	2	0.703
Continuous Performance Test	0.89	2	0.641	0.29	2	0.864
Executive Functioning	0.47	2	0.792	0.33	2	0.849
Delayed Penn Word Memory Task	0.28	2	0.870	0.18	2	0.916

Note. All models were adjusted for practice and second order age effects. Recovery adjusted models included an additional adjustment for the time since the most recent binge episode.

Figure 4.2. Association between Executive Functioning and grade point average. The fit line is a generalized additive model with 95% confidence bands.

percent of the estimates for individual Continuous Performance observations demonstrated improvement (increased scores) attributable to HED drinking exposure (Figure 4.3). All but one of the visual learning scores demonstrated decrements attributable to HED, but these estimated standardized effects were extremely small (M = -0.067, 95% CI: -0.049, 0.009). The mean effect for Continuous Performance was -0.240 (95% CI: -0.010, 0.006). Ten percent of those exposed to heavy drinking had attributable effects on Continuous Performance Scores that were greater than a tenth of a standard deviation decrease. In summary, none of the associations could be distinguished from random sampling error (p's > 0.115), and point estimates for the size of the effects were extremely small.

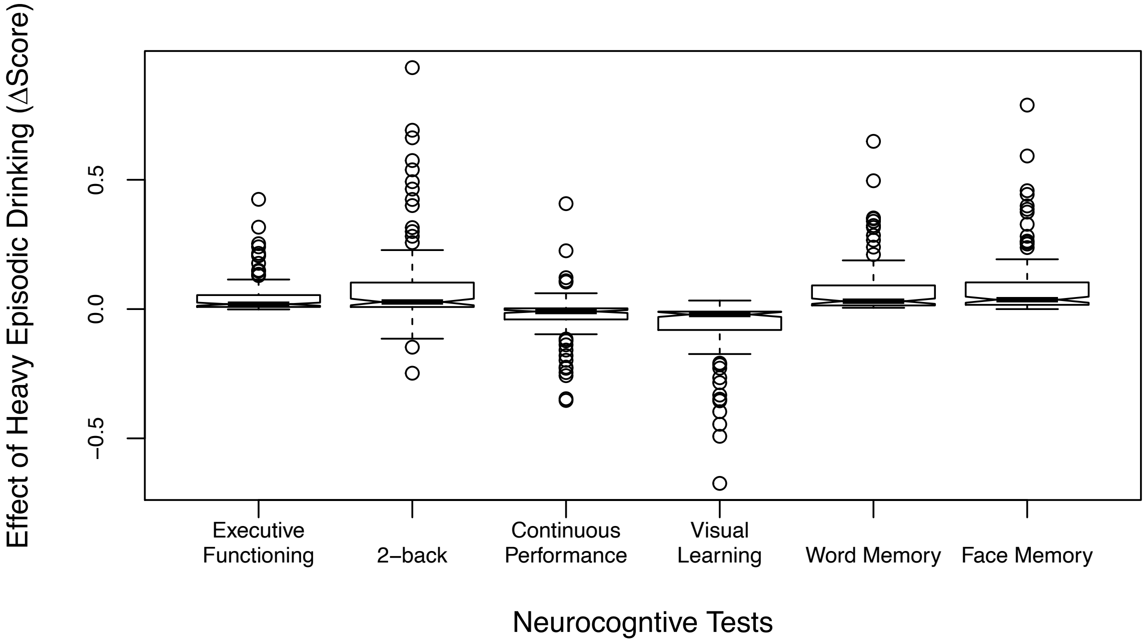


Figure 4.3. Boxplots of standardized estimated effects of heavy episodic drinking on neurocognitive test scores for periods where participants reported greater than zero heavy drinking events. Distributions of estimated changes in the neuropsychological scores that are attributable to heavy episodic drinking exposure are presented.

HED Attributable Anomalies on Neurocognition and School Grades

None of the models of the changes associated with neurocognitive anomalies attributable to heavy episodic drinking on GPA were statistically significant (likelihood ratio test $X_6^2 \leq 8.33$, p's ≥ 0.215). The strongest linear association was observed with the Continuous Performance Task where the variance explained by a Pearson product moment correlation between changes in Continuous Performance scores and changes in grade point average was extremely small (R^2 = 0.3%). Additionally, there was no evidence that the associations were masked by the linear fit; the strongest second order model (i.e., curvilinear) predicting grade point average from the attributable effects of HED on neurocognitive scores was Executive Functioning, after adjusting for age and practice effects. However, this model did not approach statistical significance (likelihood ratio test $X_6^2 = 8.33$, p = 0.215). In addition, visualization of the associations failed to identify discernable functional relationships that were not detectable by the constraints of the evaluated statistical models (Figure 4.4).

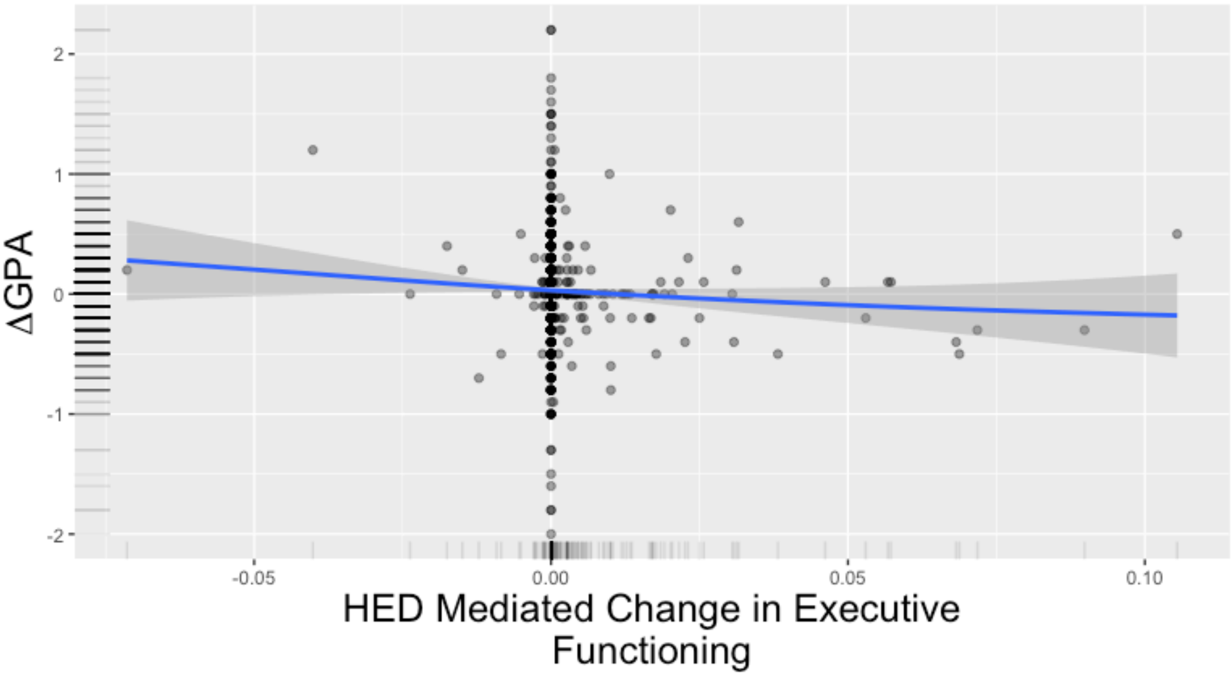


Figure 4.4. Association between prospective changes in grade point average over six months and changes in executive functioning attributable to heavy episodic drinking exposures during the prior year. The fit line is a generalized additive model with a 95% confidence band. Associations for measures of other neurocognitive domains were similar.

DISCUSSION

The current study estimated causal effects of HED on neurocognitive performance. Then the impact of these alcohol specific effects on academic performance was evaluated, by leveraging data from a large-scale longitudinal study of youth neurodevelopment. The interpretation and causal inferences benefited from the study's temporal design. The temporal precedence of the identified neurocognitive changes attributable to HED exposure to the subsequent period of scholastic activity in which academic performance was assessed was a strength of this study.

A substantial literature has firmly established a cross-sectional association between academic performance and aspects of neurocognitive functioning, most commonly general cognitive ability (Jencks, 1979; Mackintosh & Mackintosh, 2011). The strongest finding of the current study was consistent with this literature. We found correlations for unadjusted associations between each of the neuropsychological domain scores and grade point average. The strength of these correlations was lower than reported in much of the literature. Several considerations can explain this. First, some prior studies have been criticized for relying on neurocognitive tests that strongly reflect academic achievement (Kamphaus, 2005). This is particularly relevant for interpreting the early reports. More recent literature is based on neurocognitive tests that are relatively delineated from scholastic skills and experiences. The current study uses well validated and reliable measures of neurocognitive functions that are designed to be sensitive differences across the normative range of functioning and be distinct from academic achievement, and dampen socio-cultural biases (Gur et al., 2010; Gur et al., 2012). This can partially explain the apparent attenuation in the correlation. Most other studies on this topic investigate general cognitive functioning and the aligned construct of general intelligence.

In this study, only the domains of functioning that were available from the NCANDA study that were expected to be most sensitive to the effects of HED were investigated. These were attention (Tapert et al., 2002), working memory (Squeglia et al., 2012b), verbal learning (Mahmood et al., 2010; Sneider et al., 2013), and memory (Mota et al., 2013; Hanson et al., 2011; Tapert et al., 2002), and executive functioning (Mota et al., 2013). Visuospatial processing is also thought to be sensitive (Tapert & Brown, 1999), but measurement scores for the task designed to tap this cognitive function was not yet available for analysis. Cognitive functions are not completely compartmentalized and processed in delineated portions of the brain (Farah, 1994). The brain functions as a whole unit. This can help, in part, to explain the positive manifold of general cognitive ability (Jensen, 2002), as has been represented in network models of cognitive functioning (Savi, Marsman, van der Maas, & Maris, 2019; van der Maas et al., 2017). Although the functional components included in this study should be expected to correlate with general cognitive ability scores, they should be expected to be more impacted by measurement error because they are individual scores, rather than a composite score, and they may tap skills that may be of lower relative importance to overall academic performance.

After ensuring that the grades-cognitive functioning association was replicated in the current study's sample of high school students, the effects of HED on neurocognitive functioning were estimated. We employed complex curvilinear models of neurocognitive development and successively added to the developmental model terms that let individual participants' trajectories be deflected as a result of exposure to HED in the prior year. Terms that adjusted for the recency of the HED exposure were also included. This adjustment allows for neurological recovery during abstinence to be captured by the model. This adjustment is a novel strength of the current study. This allows alcohol exposure that occurred 10 months prior to an assessment to have less of an impact on neurocognitive functioning than heavy episodic drinking in the month immediately prior to testing. Standard modeling weighs binging exposure 10 months ago equally to four weeks ago. There is emerging evidence that decrements in brain functioning and neurocognitive functioning can begin to recover on a time scale of about one month during abstinence (Winward, Bekman, Hanson, Lejuez, & Brown, 2014; Brumback et al., 2015a). The modeling approach in this study was also structured so that average treatment effects (ATEs) in a causal inference framework would be ascertained. However, causal effects of HED on neurocognitive function could not be detected.

The lack of a detectable causal effect in this study does not support the conclusion that HED has no substantial effects on neurocognitive functioning. We cannot consider the current study a severe test (Mayo, 2019), because the HED exposures among the study's high school students were so low. Less than a third of the sample reported HED, overall drinking frequency was low, and the mean time since the last binge exceeded two months among bingers. Detection of true effects that are only subtle, even under strong exposures, may not be distinguishable from sampling error or can be overwhelmed by study artefacts such as practice effects. As data accumulates in the NCANDA study stronger tests will be possible, which can be combined with tools that are currently under development for adjusting for complex prior-testing effects.

Two challenges to the NCANDA study that can be addressed in future protocols are related to the measurement of academic performance and improving causal inference. The validity and reliability of grade point average as a measure of skills and knowledge is disputed (Kuncel, Credé, & Thomas, 2005). Course grades are often composite measures of student engagement and performance, where students with strong skills may have a lower grade than a student with less advanced competencies. This can happen if the weaker student turned in more assignments and attended more class sessions, in some settings. One of the most well-established threats to GPA's construct validity is known as the differential departmental grading standards (Goldman & Hewitt, 1975). Departments, instructors, and courses vary in terms of the demands of the coursework and the grading standards. There are also known threats to reliability and validity due to self-report errors, which are known to be moderated by neurocognitive functioning (Kuncel et al., 2005). We found only a moderate inter-reporter reliability in this study. Improvements in the conceptualization and measurement of academic performance can be achieved with access to grades and grade sheets at the individual course and assignment level. With digitization and standardization of gradebooks ascertainment of this data directly from participants' schools has become feasible. Reliance on richer grade data can allow for the application of item response theory-based grade point models that account for differential departmental grading standards (Young, 1990). Although future studies can be enhanced by improved measurement of academic performance, grade point average does hold some validity. There is important evidence from randomized studies that the coarse measurement provided by raw grade point averages retains valuable predictive validity for important academic and socio-economic outcomes (Cohen-Schotanus et al., 2006).

Future work can also be tailored to strengthen the causal inferences made in the current study. The causal inferences in the current study strongly rely on the assumption that all of the confounders were included in the study. Academic performance and achievement are influenced by factors beyond those measured in typical assessments of scholastic aptitude assessments. These factors often have reciprocal and interacting effects. Substance use is just one of the factors which has robust relationships with academic performance and attainment (Cox et al., 2007; Lopez-Frias & Fernandez…, 2001; Singleton & Wolfson, 2009; Friedman & Glickman…, 1985; Latvala et al., 2014; McCluskey et al., 2002; Bergen et al., 2005; Engberg & Morral, 2006; Balsa et al., 2011; Porter & Pryor, 2007; Bachman et al., 2008). The current study included developmental and demographic confounders in the causal analysis. Because the primary aims of the NCANDA project did not include unravelling the totality of causal pathways for academic performance, the potential confounders that were entered into the causal analyses were limited to components of school connectedness and attendance. Other potential confounders include scholastic attitudes and motivation (Steinmayr & Spinath, 2009), socio-economic settings (Sirin, 2005), parental involvement (Jeynes, 2005), mindset (Paunesku et al., 2015), resilience (Deb & Arora, 2012), instructor pedagogy (Freeman et al., 2014), social acceptance (Dvorsky & Langberg, 2016), behavioral engagement (Schaefer & McDermott, 1999), and negative affect (Ansary & Luthar, 2009; Hill, Locke, Lowers, & Connolly, 1999; Tomlinson et al., 2013; Kaplow et al., 2001). Many of these influences on school performance are also generated by or modulated through neurocognitive processes. Some may be dominant or keystone determinants of scholastic functioning. The current study attempted to provide a preliminary estimation of the strength of the mediated effect through neurocognitive changes due to heavy episodic drinking in adolescents. Future work based on samples with greater levels of HED exposure during high school can best advance the field if these potential confounders are included in the causal modeling and used as comparators to simultaneously gage the relative effects of these various alternative causal factors driving academic performance.

ACKNOWLEDGEMENTS

The data presented here are based on the National Consortium on Alcohol & NeuroDevelopment in Adolescence (NCANDA) Data Release NCANDA_RELEASE_4Y_REDCAP_MEASUREMENTS_V02 which was made possible through the NIH grants AA021695, AA021692, AA021696, AA021697, AA021690, AA021681, and AA021691. NCANDA data can be obtained through the National Institute of Alcohol Abuse and Alcoholism (NIAAA) at https://www.niaaa.nih.gov. The variable indexing families will be included in future releases. The authors wish to recognize .... for their helpful comments on early drafts of this manuscript.

Chapter 4 is currently being prepared for submission for publication as: Cummins, K. M., Brown, S. A., Trim, R. S., & Tapert, S. F. Mediation of heavy episodic drinking effects on academic performance through neurocognitive functioning. Kevin Cummins was the primary investigator and author of this paper.

CHAPTER 5: DISCUSSION

The body of work presented in this dissertation tackles an important research gap in the substance use literature. That gap concerns the extent that substance use, and particularly alcohol use, during adolescence results in neurocognitive changes that translate into negative academic consequences for youth (King et al., 2006). Alcohol remains the most common substance used by American high school students (Johnston et al., 2019). There is a confluence of evidence that this exposure results in changes in neuroanatomical structures, brain function, and neurocognition (Pfefferbaum et al., 2017; Spear, 2018; Lees et al., 2020; Squeglia et al., 2014; Squeglia & Gray, 2016). A solid literature also ties various causal pathways between substance use and academic functioning and performance (Bachman et al., 2008; Schulenberg et al., 1994). Although, profound effects of alcohol use among normative adolescent drinkers have not been reported, there is reason to suspect that the transient and more long-term effects of alcohol can negatively impact academic performance.

The association between academic performance and school grades is well-established (Mackintosh & Mackintosh, 2011; Jencks, 1979; Deary et al., 2007). However, there is broad suite of interacting determinants of scholastic engagement and academic performance. Indeed, behavioral compensation for reductions in neurocognitive ability may play an important role in this relationship. Future research may find that the effects of alcohol are buffered by behavior changes. It may also be that youth with greater neurocognitive performance have greater neurocognitive reserves that allow for greater compensatory plasticity. For example, work by Squeglia and colleagues (2012b) found that brain activation was substantially increased and relied upon alternative networks among HED exposed youth, yet they had similar neurocognitive functioning. The unexposed individuals may have retained greater cognitive reserves. If this is the case, we would expect that students with low baseline functioning to be most sensitive to the negative effects of HED. Some limited evidence fits with this hypothesis; for example, association between HED and decrements in grade point averages is stronger among students having been previously identified as having a learning disability (Hollar & Moore, 2004).

The current study used a representative community sample to estimate the effects of HED on academic performance (Aim 3). No association was detected (see Chapter 4). However, the evaluation did not constitute a severe test (Mayo, 2019). Drawing firm conclusions is not justified. The primary limit of this investigation is low rates of HED exposure experienced by the study's participants during the period of observation relevant to this dissertation project. Subtle effects are challenging to detect in the presence of natural variation and measurement noise. Data still being collected in the NCANDA study show an escalation of drinking among many of the study participants. The design of the study advantages research on this topic and the investigation of the effects of HED on academic performance is likely to be more definitive after additional data has been obtained.

The use of the NCANDA data for this purpose will be advantaged by the theoretical approach (Aim 1) and measurement model (Aim 2) developed in this dissertation. The factor modeling conducted in Chapter 3 provides a measurement model for one of the important outcomes of HED exposure. It also correlates of academic performance. This was executive functioning. Although the model fit only reached moderate levels acceptable features and levels of invariance were established.

The posture toward using paratheoretical framing and building of phenomenological models can assist future researchers to focus on the critical components of the conceptual devices used to represent the processes determining academic performance. This can reduce the potential for the blinding effects of theory from restricting the consideration of and comparisons with alternative determinants and causal pathways beyond neurocognitive mediation. The approach was developed in Chapter 2 and applied in Chapters 3 and 4. It will also help future investigators recognize that when they construct models of this system as fictive representations, that they decide how best to balance realism, precision, and generalizability to best match the needs of the particular investigation. This can include a subjective element and is arguably the second most important point at which induvial scientists can make unique contributions (the other being the generation of question-hypothesis couplets). Models of determinants of academic performance certainly will benefit from breaking down the disciplinary silos that currently characterize this topic of investigation. Breaking free from disciplinary conventions for theory and moving to phenomenological models (see Chapter 2) is likely to lead investigators to recognize that definitive conclusions requires simultaneous consideration of the complex and interacting sets of determinants, which are almost always studied in isolation in contemporary research. NCANDA has access to measures and surrogates for some of the determinants such as school connectedness and truancy. Supplemental data on determinants such as parental support and classroom instructional practices would benefit this effort. In the absence of supplemental data, results of future analyses can still remain highly valuable in the parameterization of determinant rich phenomenological models that synthesize and integrate studies of various determinants.

REFERENCES

Abbott, I. (1980). Theories dealing with the ecology of landbirds on islands. *Advances in Ecological Research*, *11*, 329-371.

Achinstein, P. (1968). *Concepts of Science: A Philosophical Analysis*. Baltimore: The Johns Hopkins Press.

Achterberg, C., & Miller, C. (2004). Is one theory better than another in nutrition education? a viewpoint: more is better. *Journal of Nutrition Education and Behavior*, *36*(1), 40-42.

Ahlström, S. K., & Österberg, E. L. (2004). International perspectives on adolescent and young adult drinking. *Alcohol Research*, *28*(4), 258.

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: the worldviews of AIC and BIC. *Ecology*, *95*(3), 631-636.

Ajzen, I. (2011). The theory of planned behavior. In P. A. van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology*. New York: Lawrence Erlbaum Associates.

Akçakaya, H. R., Franklin, J., Syphard, A. D., & Stephenson, J. R. (2005). Viability of Bell's sage sparrow (Amphispiza belli ssp. belli): altered fire regimes. *Ecological Applications*, *15*(2), 521-531.

Akshoomoff, N., Newman, E., Thompson, W. K., McCabe, C., Bloss, C. S., Chang, L. et al. (2014). The NIH Toolbox Cognition Battery: Results from a large normative developmental sample (PING). *Neuropsychology*, *28*(1), 1.

Albada, A., Ausems, M., Bensing, J. M., & van Dulmen, S. (2009). Tailored information about cancer risk and screening: a systematic review. *Patient Education and Counseling*, *77*(2), 155-171.

Allen, T. F. H., & Starr, T. B. (2017). *Hierarchy*. Chicago: University of Chicago Press.

American Association for the Advancement of Science. (2005). Evolution on the Front Line: Q & A on Evolution and Intelligent Design. Retrieved October 29, 2005, https://www.aaas.org/news/press_room/evolution/qanda.shtml.

Ammerman, A. S., Lindquist, C. H., Lohr, K. N., & Hersey, J. (2002). The efficacy of behavioral interventions to modify dietary fat and fruit and vegetable intake: a review of the evidence. *Preventive Medicine*, *35*(1), 25-41.

Anonymous. (2014). “The NIH Toolbox Cognition Battery: Results from a large normative developmental sample (PING)”: Correction to Akshoomoff et al. (2013) (2014). Neuropsychology, 28(2), 319. *Neuropsychology*, *28*(2), 319.

Ansary, N. S., & Luthar, S. S. (2009). Distress and academic achievement among adolescents of affluence: A study of externalizing and internalizing problem behaviors and school performance. *Development and psychopathology*.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3-25.

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495-508.

Austin, G., Polik, J., Hanson, T., & Zheng, C. (2018). *School climate, substance use, and student well-being in California, 2015-2017*. San Fransisco: WestEd.

Bachman, J. G., Johnston, L. D., O’Malley, P. M., Schulenberg, J. E., & Miech, R. A. (2015). The Monitoring the Future project after four decades: Design and procedures.

Bachman, J. G., Schulenberg, J. E., Freedman-Doan, P., Messersmith, E. E., O’Malley, P. M., & Johnston, L. D. (2008). *The Education-Drug Use Connection: How Successes and Failures in School Relate to Adolescent Smoking, Drinking, Drug Use, and Delinquency*. New York, New York: Lawrence Erlbaum Associates.

Balsa, A. I., Giuliano, L. M., & French, M. T. (2011). The effects of alcohol use on academic achievement in high school. *Economics of Education Review*.

Banca, P., Lange, I., Worbe, Y., Howell, N. A., Irvine, M., Harrison, N. A. et al. (2016). Reflection impulsivity in binge drinking: behavioural and volumetric correlates. *Addict Biol*, *21*(2), 504-515.

Bandura, A. (1986). *Social Foundations of Thought and Action: a Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bartholomew Eldredge, L. K., Markham, C. M., Ruiter, R. A. C., Fernández, M. E., Kok, G., & Parcel, G. S. (2016). *Planning health promotion programs: an intervention mapping approach*. John Wiley & Sons.

Bartholomew, L. K., & Mullen, P. D. (2011). Five roles for using theory and evidence in the design and testing of behavior change interventions. *Journal of Public Health Dentistry*, *71*, S20-S33.

Bava, S., Jacobus, J., Mahmood, O., Yang, T. T., & Tapert, S. F. (2010). Neurocognitive correlates of white matter quality in adolescent substance users. *Brain and cognition*, *72*(3), 347-354.

Beamish, R. J., & Rothschild, B. J. (2009). *The future of fisheries science in North America* (31). Springer Science & Business Media.

Bergen, H. A., Martin, G., Roeger, L., & Allison, S. (2005). Perceived academic performance and alcohol, tobacco and marijuana use: Longitudinal relationships in young community adolescents. *Addictive behaviors*.

Billor, N., Hadi, A. S., & Velleman, P. F. (2000). BACON: blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis*, *34*(3), 279-298.

Bollen, K. A. (2014). *Structural Equations with Latent Variables*. John Wiley & Sons.

Boniolo, G. (2007). *On Scientific Representations: From Kant to a New Philosophy of Science*. Springer.

Boniolo, G. (1997). On a unified theory of models and thought experiments in natural sciences. *International Studies in the Philosophy of Science*, *11*(2), 121-142.

Boniolo, G., Petrovich, C., & Pisent, G. (2002). Notes on the philosophical status of nuclear physics. *Foundations of Science*, *7*(4), 425-452.

Boonstra, R., Krebs, C. J., & Stenseth, N. C. (1998). Population cycles in small mammals: the problem of explaining the low phase. *Ecology*, *79*(5), 1479-1488.

Braysher, M., Buckmaster, T., Saunders, G., & Krebs, C. J. (2012). Principles underpinning best practice management of the damage due to pests in Australia. In *25(25)* (pp. 300-307). Monterey, California: University of California, Davis.

Brown, S. A., Brumback, T., Tomlinson, K., Cummins, K., Thompson, W. K., Nagel, B. J. et al. (2015). The National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA): a multisite study of adolescent development and substance use. *Journal of Studies on Alcohol and Drugs*, *76*(6), 895-908.

Brown, S. A., Myers, M. G., Lippke, L., Tapert, S. F., Stewart, D. G., & Vik, P. W. (1998). Psychometric evaluation of the Customary Drinking and Drug Use Record (CDDR): a measure of adolescent alcohol and drug involvement. *Journal of Studies on Alcohol*, *59*(4), 427-438.

Brown, S. A., McGue, M., Maggs, J., Schulenberg, J., Hingson, R., Swartzwelder, S. et al. (2008). A Developmental Perspective on Alcohol and Youths 16 to 20 Years of Age. *Pediatrics*, *121*(Supplement 4), S290-S310.

Brown, T. A. (2015). Confirmatory Factor Analysis for Applied Research, Second Edition.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-136). Newbury Park: Sage.

Brumback, T., Squeglia, L. M., Jacobus, J., Pulido, C., Tapert, S. F., & Brown, S. A. (2015a). Adolescent heavy drinkers' amplified brain responses to alcohol cues decrease over one month of abstinence. *Addictive behaviors*, *46*, 45-52.

Brumback, T., Squeglia, L. M., Jacobus, J., Pulido, C., Tapert, S. F., & Brown, S. A. (2015b). Adolescent heavy drinkers' amplified brain responses to alcohol cues decrease over one month of abstinence. *Addictive Behaviors*, *46*, 45-52.

Brumback, T., Thompson, W., Cummins, K., Brown, B., & Tapert, S. (2020). Psychosocial Predictors of Substance Use in Adolescents and Young Adults: Longitudinal Risk and Protective Factors. *Addictive Behaviors*, *Under Review*.

Bruno, J. F., Stachowicz, J. J., & Bertness, M. D. (2003). Inclusion of facilitation into ecological theory. *Trends in Ecology & Evolution*, *18*(3), 119-125.

Bryant, A. L., Schulenberg, J. E., & O'malley…, P. M. (2003). How academic achievement, attitudes, and behaviors relate to the course of substance use during adolescence: A 6-year, multiwave national longitudinal study. *Journal of research …*.

Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer Science & Business Media.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23-35.

Cadigan, J. M., Haeny, A. M., Martens, M. P., Weaver, C. C., Takamatsu, S. K., & Arterberry, B. J. (2015). Personalized drinking feedback: a meta-analysis of in-person versus computer-delivered interventions. *Journal of Consulting and Clinical Psychology*, *83*(2), 430.

Callaway, R. M. (1997). Positive interactions in plant communities and the individualistic-continuum concept. *Oecologia*, *112*(2), 143-149.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D. et al. (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, *321*(7262), 694-696.

Carey, K. B., Scott-Sheldon, L. A., Carey, M. P., & DeMartini, K. S. (2007). Individual-level interventions to reduce college student drinking: a meta-analytic review. *Addictive Behavior*, *32*(11), 2469-2494.

Cepeda, J. A., Eritsyan, K., Vickerman, P., Lyubimova, A., Shegay, M., Odinokova, V. et al. (2018). Potential impact of implementing and scaling up harm reduction and antiretroviral therapy on HIV prevalence and mortality and overdose deaths among people who inject drugs in two Russian cities: a modeling study. *The Lancet HIV*, *5*(10), e578-e587.

Chaiken, S., & Trope, Y. (1999). *Dual-process Theories in Social Psychology*. New York: Guilford Press.

Chalmers, A. (2010). The life of theory in the new experimentalism. In D. G. Mayo & A. Spanos (Eds.), *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliablity, and the Objectivity and Rationality of Science* (pp. 58-72). New York: Cambridge University Press.

Chalmers, A. F. (2013). *What Is This Thing Called Science?: Forth Edition*. Hackett Publishing.

Anderson, D. R., & Burnham, K. (2004). Model selection and multi-model inference. *Second. NY: Springer-Verlag*, 63.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504.

Chesson, P. L., & Case, T. J. (1986). Overview: nonequilibrium community theories: chance, variability, history. In J. Diamond & T. J. Case (Eds.), *Community Ecology* (pp. 229-239). New York: Harper and Row.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Chitty, D. (1996). *Do Lemmings Commit Suicide: Beautiful Hypotheses and Ugly Facts*. Oxford: Oxford University Press.

Christiansen, F. B., & Fenchel, T. M. (2012). *Theories of Populations in Biological Communities*. Springer Science & Business Media.

Chung, T., & Jackson, K. M. (2019). Adolescent alcohol use. In R. A. Zucker & S. A. Brown (Eds.), *The Oxford Handbook of Adolescent Substance Abuse* (p. 131). New York, New York: Oxford University Press.

Cohen-Schotanus, J., Muijtjens, A. M., Reinders, J. J., Agsteribbe, J., van Rossum, H. J., & van der Vleuten, C. P. (2006). The predictive validity of grade point average scores in a partial lottery medical school admission system. *Med Educ*, *40*(10), 1012-1019.

Cohen, D. A., & Rice, J. (1997). Parenting styles, adolescent substance use, and academic achievement. *Journal of Drug Education*, *27*(2), 199-211.

Collie, J. S., Richardson, K., & Steele, J. H. (2004). Regime shifts: can ecological theory illuminate the mechanisms. *Progress in Oceanography*, *60*(2-4), 281-302.

Converse, S. J., & Moore, C. T. (2013). Demographics of reintroduced populations: estimation, modeling, and decision analysis. *Journal of Wildlife Management*, *77*(6), 1081-1093.

Cox, R. G., Zhang, L., & Johnson…, W. D. (2007). Academic performance and substance use: findings from a state survey of public high school students. *Journal of School …*.

Crews, F. T., Vetreno, R. P., & Broadwater…, M. A. (2016). Adolescent Alcohol Exposure Persistently Impacts Adult Neurobiology and Behavior. *Pharmacological Review*, *68*, 1074-1109.

Crosby, R. A., Kegler, M. C., & DiClemente, R. J. (2009). Theory in health promotion practice and research. In *Emerging Theories in Health Promotion Practice and Research* (pp. 3-18). San Francisco: Jossey-Bass.

Crum, R. M., Juon, H. S., Green, K. M., & Robertson…, J. (2006). Educational achievement and early school behavior as predictors of alcohol-use disorders: 35-year follow-up of the Woodlawn Study. *... of studies on alcohol*.

Cummins, K. M., Pitpitan, E. V., Reed, M., & Zuniga, M. (Under Review). Reforming the use of theory in health promotion: lessons from ecology. *Health Pscyhology Reviews*.

Cummins, K. M., & Marks, C. L. (2020). Farewell to Bright-Line: A Guide to Reporting Quantitative Results Without the S-Word. *Frontiers in Psychology*, *In Press*.

Cummins, K. M., Pitpitan, E. V., Brown, S. A., Clark, D. B., Brumback, T., Trim, R. S. et al. (2020). Comparison of factor analysis models applied to the NCANDA neuropsychological test battery. *Journal of Clinical and Experimental Neuropsychology*, *Under Review*.

Datta, J., & Petticrew, M. (2013). Challenges to evaluating complex interventions: a content analysis of published papers. *BMC Public Health*, *13*, 568.

Davis, R., Campbell, R., Hildon, Z., Hobbs, L., & Michie, S. (2015). Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychology Review*, *9*(3), 323-344.

Dayton, P. (1980). This week's citation classic. *Current Contents*, *32*, 156.

De Wit, R. (2005). Do all ecosystems maximise their distance with respect to thermodynamic equilibrium? a comment on the ecological law of thermodynamics (ELT), proposed by Sven Erik Jørgensen. *Scientia Marina*, *69*(3), 427-434.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13-21.

Deb, A., & Arora, M. (2012). Resilience and academic achievement among adolescents. *Journal-Indian Academy of Applied Psychology*, *38*(1), 93-101.

Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society*, *9*(6), 936-946.

DiClemente, R. J., Crosby, R. A., & Kegler, M. C. (2009). *Emerging Theories in Health Promotion Practice and Research*. John Wiley & Sons.

Dochtermann, N. A., & Jenkins, S. H. (2011). Developing multiple hypotheses in behavioral ecology. *Behavioral Ecology and Sociobiology*, *65*(1), 37-45.

Dvorsky, M. R., & Langberg, J. M. (2016). A review of factors that promote resilience in youth with ADHD and ADHD symptoms. *Clinical child and family psychology review*, *19*(4), 368-391.

Engberg, J., & Morral, A. R. (2006). Reducing substance use improves adolescents' school attendance. *Addiction*, *101*(12), 1741-1751.

Estabrook, R., & Neale, M. (2013). A comparison of factor score estimation methods in the presence of missing data: Reliability and an application to nicotine dependence. *Multivariate behavioral research*, *48*(1), 1-27.

Farah, M. J. (1994). Neuropsychological inference with an interactive brain: A critique of the "locality" assumption. *Behavioral and Brain Sciences*, *17*(1), 43-61.

Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research*, *63*(3), 324-327.

Fishbein, M., & Yzer, M. C. (2003). Using theory to design effective health behavior interventions. *Communication theory*, *13*(2), 164-183.

Flyvbjerg, B. (2001). *Making social science matter: Why social inquiry fails and how it can succeed again*. Cambridge, UK: Cambridge University Press.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H. et al. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415.

Fregly, A. R., Graybiel, A., & Smith, M. J. (1972). Walk on floor eyes closed (WOFEC): a new addition to an ataxia test battery. *Aerospace Medicine*, *43*(4), 395-399.

Friedman, A. S., & Glickman…, N. (1985). Does drug and alcohol use lead to failure to graduate from high school. *Journal of Drug …*.

Frigg, R., & Hartmann, S. (2006). Scientific Models. In S. Sarkar & J. Pfeifer (Eds.), *The Philosophy of Science: An Encyclopedia* (pp. 740-749). London: Routledge.

Galaif, E. R., Sussman, S., Newcomb, M. D., & Locke, T. F. (2007). Suicidality, depression, and alcohol use among adolescents: a review of empirical findings. *International journal of adolescent medicine and health*, *19*(1), 27-36.

Galbraith, S., Bowden, J., & Mander, A. (2017). Accelerated longitudinal designs: An overview of modeling, power, costs and handling missing data. *Statistical Methods Medical Research*, *26*(1), 374-398.

Gardner, B., Wardle, J., Poston, L., & Croker, H. (2011). Changing diet and physical activity to reduce gestational weight gain: a meta-analysis. *Obesity Reviews*, *12*(7), e602-e620.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Genderson, M. R., Dickinson, D., Diaz-Asper, C. M., Egan, M. F., Weinberger, D. R., & Goldberg, T. E. (2007). Factor analysis of neurocognitive tests in a large sample of schizophrenic probands, their siblings, and healthy controls. *Schizophrenia Research*, *94*(1-3), 231-239.

Giancola, P. R., & Mezzich, A. C. (2000). Neuropsychological deficits in female adolescents with a substance use disorder: better accounted for by conduct disorder. *J Stud Alcohol*, *61*(6), 809-817.

Glanz, K., & Bishop, D. B. (2010). The role of behavioral science theory in development and implementation of public health interventions. *Annual Review of Public Health*, *31*, 399-418.

Glanz, K., Rimer, B. K., & Viswanath, K. (2015). *Health Behavior: Theory, Research, and Practice* (Fifth Edition ed.). San Fransisco: Jossey-Bass.

Glanz, K. (2005). *Theory at a Glance: a Guide for Health Promotion Practice. Second Edition.* (2nd ed. ed.). Bethesda: U.S. Dept. of Health and Human Services, National Cancer Institute.

Goldman, R. D., & Hewitt, B. N. (1975). Adaptation-level as an explanation for differential standards in college grading. *Journal of Educational Measurement*, 149-161.

Goldstein, B. I. (2009). Subthreshold Substance Use and the Treatment of Resistant Depression in Adolescents. *Child & Adolescent Psychiatry*, *48*(12), 1182-1192.

Goodson, P. (2010). *Theory in Health Promotion Research and Practice: Thinking Outside the Box*. Sudbury, Mass: Jones and Bartlett.

Gore, F. M., Bloem, P. J., Patton, G. C., Ferguson, J., Joseph, V., Coffey, C. et al. (2011). Global burden of disease in young people aged 10-24 years: a systematic analysis. *Lancet*, *377*(9783), 2093-2102.

Gorelick, R. (2011). What is theory? *Ideas in Ecology and Evolution*, *4*, 1-10.

Grant, B. F., & Dawson, D. A. (1997). Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: results from the National Longitudinal Alcohol Epidemiologic Survey. *Journal of substance abuse*, *9*, 103-110.

Grant, J. D., Scherrer, J. F., Lynskey, M. T., Lyons, M. J., Eisen, S. A., Tsuang, M. T. et al. (2006). Adolescent alcohol use is a risk factor for adult alcohol and drug dependence: evidence from a twin design. *Psychological Medicine*, *36*(1), 109-118.

Gray, K. (2017). How to map theory: reliable methods are fruitless without rigorous theory. *Perspectives on Psychological Science*, *12*(5), 731-741.

Graybiel, A., Smith, C. R., Guedry Jr, F. E., Miller, E. F., Fregly, A. R., & Cramer, D. B. (1972). Idiopathic progressive vestibular degeneration. *Annals of Otology, Rhinology & Laryngology*, *81*(2), 165-178.

Greve, W. (2001). Traps and gaps in action explanation: Theoretical problems of a psychology of human action. *Psychological Review*, *108*(2), 435.

Griffin, W. A. (2006). Agent-based modeling for the theoretical biologist. *Biological Theory*, *1*(4), 404-409.

Grime, J. P. (2007). Plant strategy theories: a comment on Craine (2005). *Journal of Ecology*, *95*(2), 227-230.

Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B. et al. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. *Journal of Neuroscience Methods*, *187*(2), 254-262.

Gur, R. C., Richard, J., Calkins, M. E., Chiavacci, R., Hansen, J. A., Bilker, W. B. et al. (2012). Age group and sex differences in performance on a computerized neurocognitive battery in children age 8− 21. *Neuropsychology*, *26*(2), 251.

Hacking, I. (1983). *Representing and Intervening*. Cambridge University Press.

Hagger, M. S., & Weed, M. (2019). DEBATE: Do interventions based on behavioral theory work in the real world. *Int J Behav Nutr Phys Act*, *16*(1), 36.

Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, *10*(4), 371-388.

Haller, B. C. (2014). Theoretical and empirical perspectives in ecology and evolution: a survey. *BioScience*, *64*(10), 907-916.

Haller, M., Norman, S. B., Cummins, K., Trim, R. S., Xu, X., Cui, R. et al. (2016). Integrated cognitive behavioral therapy versus cognitive processing therapy for adults with depression, substance use disorder, and trauma. *Journal of Substance Abuse Treatment*, *62*, 38-48.

Halstead, W. C. (1947). *Brain and intelligence; a quantitative study of the frontal lobes.* Chicago: University of Chicago Press.

Hanson, K. L., Cummins, K., Tapert, S. F., & Brown, S. A. (2011). Changes in neuropsychological functioning over 10 years following adolescent substance abuse treatment. *Psychology of Addictive Behaviors*, *25*(1), 127-134.

Harper, J. L. (1980). Plant demography and ecological theory. *Oikos*, *35*(2), 244-253.

Harré, R. (2004). *Modeling: Gateway to the Unknown: a Work* (1). New York: Elsevier.

Harrell, F. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. New York: Springer.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models* (43). CRC press.

Hawe, P. (2015). Lessons from complex interventions to improve health. *Annual Review of Public Health*, *36*, 307-323.

Hawking, S. (2011). *The Grand Design*. New York: Random House Digital, Inc.

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: implications for substance abuse prevention. *Psychological bulletin*, *112*(1), 64.

Hayduk, L. (2014). Seeing Perfectly Fitting Factor Models That Are Causally Misspecified. *Educational and Psychological Measurement*, *74*(6), 905-926.

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models. *BMC Medical Research Methodology*, *12*(1), 159.

Hilborn, R., & Ovando, D. (2014). Reflections on the success of traditional fisheries management. *ICES Journal of Marine Science*, *71*(5), 1040-1046.

Hill, S. Y., Locke, J., Lowers, L., & Connolly, J. (1999). Psychopathology and achievement in children at high risk for developing alcoholism. ... *of the American Academy of Child ….*

Hodges, K. E. (2008). Defining the problem: terminology and progress in ecology. *Frontiers in Ecology and the Environment*, *6*(1), 35-42.

Hollar, D., & Moore, D. (2004). Relationship of substance use by students with disabilities to long-term educational, employment, and social outcomes. *Substance Use & Misuse*, *39*(6), 931-962.

Homer, J. (1996). Why we iterate: scientific modeling in theory and practice. *Systems Dynamics Review*, *12*(1), 1-19.

Homer, J., & Hirsch, S. (2006). System dynamics modeling for public health: background and opportunities. *American Journal of Public Health*, *96*(3), 452-458.

Hu, L., & Bentler, P. M. (1999a). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, *6*(1), 1-55.

Hu, L., & Bentler, P. M. (1999b). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, *6*(1), 1-55.

Huang, F. L., & Cornell, D. G. (2016). Using Multilevel Factor Analysis With Clustered Data. *Journal of Psychoeducational Assessment*, *34*(1), 3-14.

Hurlbert, S. H., Levine, R. A., & Utts, J. (2019). Coup de Grâce for a Tough Old Bull: "Statistically Significant" Expires. *The American Statistician*, *73*(sup1), 352-357.

Cody, M., & Diamond, J. (Eds.). (1975). *Variations on a theme by Robert MacArthur*. Harvard University Press.

Imenda, S. (2014). Is there a conceptual difference between theoretical and conceptual frameworks. *Journal of Social Sciences*, *38*(2), 185-195.

J. Conrod, P., & Nikolaou, K. (2016). Annual Research Review: On the developmental neuropsychology of substance use disorders. *Journal of Child Psychology and Psychiatry*, *57*(3), 371-394.

Jacobus, J., & Tapert, S. F. (2013). Neurotoxic Effects of Alcohol in Adolescence. *Annual Review of Clinical Psychology*, *9*(1), 703-721.

James, J. Y., Weinberger, R., Moore, T. M., Calkins, M. E., Guri, Y., McDonald-McGinn, D. M. et al. (2016). Performance on a computerized neurocognitive battery in 22q11. 2

deletion syndrome: A comparison between US and Israeli cohorts. *Brain and cognition*, *106*, 33-41.

Jencks, C. (1979). *Who gets ahead? The determinants of economic success in America.* New York, New York: Basic Books.

Jensen, A. R. (2002). Psychometric g: Definition and substantiation. In *The General Factor of Intelligence* (pp. 51-66). Mahwah, New Jersey: Psychology Press.

Jernigan, T. L., Brown, S. A., & Dowling, G. J. (2018). The adolescent brain cognitive development study. *Journal of Research on Adolescence*, *28*(1), 154-156.

Jeynes, W. H. (2005). A meta-analysis of the relation of parental involvement to urban elementary school student academic achievement. *Urban education*, *40*(3), 237-269.

Johnston, L. D., Miech, R. A., O'Malley, P. M., Bachman, J. G., Schulenberg, J. E., & Patrick, M. E. (2018). *Monitoring the Future national survey results on drug use, 1975-2017: Overview, key findings on adolescent drug use*. Ann Arbor, Michigan: Institute for Social Research, The University of Michigan.

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*.

Jørgensen, S. E., Fath, B. D., Bastianoni, S., Marques, J. C., Muller, F., Nielsen, S. N. et al. (2011). *A New Ecology: Systems Perspective*. New York: Elsevier.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kamphaus, R. W. (2005). *Clinical assessment of child and adolescent intelligence*. Springer Science+ Business Media.

Kaplan, R., & Saccuzzo, D. (2008). *Psychological Testing: Principles, Applications, and Issues*. Cengage Learning.

Kaplow, J. B., Curran, P. J., Angold, A., & Costello, E. J. (2001). The prospective relation between dimensions of anxiety and the initiation of adolescent alcohol use. *Journal of clinical child psychology*, *30*(3), 316-326.

Kim, E. S., & Willson, V. L. (2014). Testing Measurement Invariance Across Groups in Longitudinal Data: Multigroup Second-Order Latent Growth Model. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 566-576.

Kim, N. (1997).

King, K. M., Meehan, B. T., Trim, R. S., & Chassin, L. (2006). Substance use and academic outcomes: Synthesizing findings and future directions. *Addiction*, *101*(12), 1688-1689.

Kiviniemi, M. T., Ellis, E. M., Hall, M. G., Moss, J. L., Lillie, S. E., Brewer, N. T. et al. (2018). Mediation, moderation, and context: Understanding complex relations among cognition, affect, and health behaviour. *Psychology and Health*, *33*(1), 98-116.

Kolasa, J. (2011). Theory makes ecology evolve. In S. Scheiner & M. Willig (Eds.), *The Theory of Ecology* (pp. 21-50). Chicago: University of Chicago Press.

Koob, G. F., & Volkow, N. D. (2016). Neurobiology of addiction: a neurocircuitry analysis. *The Lancet Psychiatry*, *3*(8), 760-773.

Krebs, C. J. (2017). 10 possible limitations of current ecological theory. *Australian Zoologist*, *38*(3), 267-271.

Krieger, N., & Zierler, S. (1996). What explains the public's health?: a call for epidemiologic theory. *Epidemiology*, 107-109.

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research*, *33*(2), 188-229.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of educational research*, *75*(1), 63-82.

Lakatos, I. (1968). II-criticism and the methodology of scientific research programmes. *Proceedings of the Aristotelian Society*, *69*, 149-186.

Lambert, K., & Brittan, G. G. (1992). *An Introduction to the Philosophy of Science: Forth Edition*. Atascadero, CA: Ridgeview Publishing Company.

Latvala, A., Rose, R. J., Pulkkinen, L., Dick, D. M., Korhonen, T., & Kaprio, J. (2014). Drinking, smoking, and educational achievement: cross-lagged associations from adolescence to adulthood. *Drug Alcohol Depend*, *137*, 106-113.

Lawson, D. M., Regan, H. M., Zedler, P. H., & Franklin, J. (2012). Cumulative effects of land use, altered fire regime and climate change on persistence of ceanothus verrucosus, a rare, fire-dependent plant species. *Global Change Biology*, *18*(9), 2980-2980.

Le Berre, A.-P. (2019). Alcohol use disorder: permanent and transient effects on the brain and neuropsychological functions. In M. L. Alosco & R. A. Stern (Eds.), *The Oxford Handbook of Adult Cognitive Disorders* (pp. 307-337). New York, NY: Oxford University Press.

Le Berre, A.-P., Fama, R., & Sullivan, E. V. (2017). Executive Functions, Memory, and Social Cognitive Deficits and Recovery in Chronic Alcoholism: A Critical Review to

Inform Future Research. *Alcoholism: Clinical and Experimental Research*, *41*(8), 1432-1443.

Lees, B., Meredith, L. R., Kirkland, A. E., Bryant, B. E., & Squeglia, L. M. (2020). Effect of alcohol use on the adolescent brain and behavior. *Pharmacology Biochemistry and Behavior*, 172906.

Lehman, C. L., & Tilman, D. (1997). Competition in spatial habitats. *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*, 185-203.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual review of psychology*, *66*, 799-823.

Lerner, J. S., & Shonk, K. (2010). How anger poisons decision making. *Harv Bus Rev*, *88*(9), 26.

Levin, S. A. (1992). The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, *73*(6), 1943-1967.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*(4), 421-431.

Levy, D. T., Bauer, J. E., & Lee, H. R. (2006). Simulation modeling and tobacco control: creating more robust public health policies. *American Journal of Public Health*, *96*(3), 494-498.

Lewin, K. (1951). *Field Theory in Social Science*. New York: Harper.

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*. Oxford University Press, USA.

Lippke, S., & Ziegelmann, J. P. (2008). Theory-based health behavior change: developing, testing, and applying theories for evidence-based interventions. *Applied Psychology*, *57*(4), 698-716.

Little, R. J., & Rubin, D. B. (1986). *Statistical Analysis with Missing Data*. New York, NY: John Wiley & Sons, Inc.

Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. New York, NY: Guilford press.

Lock, M., Post, D., Dollman, J., & Parfitt, G. (2020). Efficacy of theory-informed workplace physical activity interventions: a systematic literature review with meta-analyses. *Health Psychol Rev*, 1-25.

Lopez-Frias, M., & Fernandez…, M. D. L. F. (2001). Alcohol consumption and academic performance in a population of Spanish high school students. *... of studies on alcohol*.

Lucas, K., & Lloyd, B. (2005). *Health Promotion*. Thousand Oaks, CA: SAGE.

MacArthur, R. (1962). Growth and regulation of animal populations. *Ecology*, *43*(3), 579.

Mackintosh, N., & Mackintosh, N. J. (2011). *IQ and human intelligence*. Oxford University Press.

Mahmood, O. M., Jacobus, J., Bava, S., Scarlett, A., & Tapert, S. F. (2010). Learning and memory performances in adolescent users of alcohol and marijuana: interactive effects. *Journal of Studies on Alcohol and Drugs*, *71*(6), 885-894.

Margalef, R. (1963). On certain unifying principles in ecology. *The American Naturalist*, *97*(897), 357-374.

Marquet, P. A., Allen, A. P., Brown, J. H., Dunne, J. A., Enquist, B. J., Gillooly, J. F. et al. (2014). On theory in ecology. *BioScience*, *64*(8), 701-710.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B. et al. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23*(3), 524.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S. et al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural equation modeling: A multidisciplinary journal*, *16*(3), 439-476.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) …. *Structural equation modeling*.

Martin, A. J. (2011). Holding back and holding behind: grade retention and students' non-academic and academic outcomes. *British Educational Research Journal*, *37*(5), 739-763.

Masterson, C. J., Tuttle, J., & Maerlender, A. (2019). Confirmatory factor analysis of two computerized neuropsychological test batteries: Immediate post-concussion assessment and cognitive test (ImPACT) and C3 logix. *Journal of Clinical and Experimental Neuropsychology*, *41*(9), 925-932.

Matthewson, J. (2011). Trade-offs in model-building: A more target-oriented approach. *Studies in History and Philosophy of Science Part A*, *42*(2), 324-333.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, *261*(5560), 459.

May, R. M. (1981). The role of theory in ecology. *American Zoologist*, *21*(4), 903-910.

May, R. M., & Oster, G. F. (1976). Bifurcations and dynamic complexity in simple ecological models. *The American Naturalist*, *110*(974), 573-599.

Mayo, D. G. (1991). Novel evidence and severe tests. *Philosophy of Science*, *58*(4), 523-552.

Mayo, D. G. (2019). *Statistical Inference as Severe Testing*. Cambridge University Press, Cambridge.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine*, *32*(19), 3388-3414.

McCain, K. W. (2015). Nothing as practical as a good theory: does Lewin's Maxim still have salience in the applied social sciences. *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, *52*(1), 1-4.

McCluskey, C. P., Krohn, M. D., & Lizotte…, A. J. (2002). Early substance use and school achievement: An examination of Latino, White, and African American youth. *Journal of Drug ….*

McIntosh, R. P. (1980). The background and some current problems of theoretical ecology. *Synthese*, *43*(2), 195-255.

McIntosh, R. P. (1986). *The Background of Ecology*. Cambridge University Press.

McMullin, E. (1976). The fertility of theory and the unit for appraisal in science. In *Essays in Memory of Imre Lakatos* (pp. 395-432). New York: Springer.

Meehl, P. E. (1990). Appraising and amending theories: the strategy of lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*(2), 108-141.

Meier, P. S., Barrowclough, C., & Donmall, M. C. (2005). The role of the therapeutic alliance in the treatment of substance misuse: a critical review of the literature. *Addiction*, *100*(3), 304-316.

Mendl, M., & Paul, E. S. (2020). Animal affect and decision-making. *Neurosci Biobehav Rev*, *112*, 144-163.

Michie, S. F., West, R., Campbell, R., Brown, J., & Gainforth, H. (2014). *ABC of behaviour change theories*. Silverback Publishing.

Michie, S., Jochelson, K., Markham, W. A., & Bridle, C. (2009). Low income groups and behaviour change interventions: a review of intervention content, effectiveness and theoretical frameworks. *Journal of Epidemiology & Community Health*, *63*, 610-622.

Michie, S., Johnston, M., Francis, J., Hardeman, W., & Eccles, M. (2008). From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. *Applied Psychology*, *57*(4), 660-680.

Michie, S., & Prestwich, A. (2010). Are interventions theory-based? development of a theory coding scheme. *Health Psychology*, *29*(1), 1.

Miech, R. A., Schulenberg, J. E., Johnston, L. D., Bachman, J. G., O'Malley, P. M., & Patrick, M. E. (2017). National Adolescent Drug Trends in 2017: Findings Released. Retrieved April 29, 2018, from http://www.monitoringthefuture.org.

Miller, A. L., Lo, S. L., Bauer, K. W., & Fredericks, E. M. (2020). Developmentally informed behaviour change techniques to enhance self-regulation in a health promotion context: a conceptual review. *Health Psychology Review*, *14*(1), 116-131.

Minsky, M. (1965). Matter, mind and models.

Miyazaki, Y., & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods*, *5*(1), 44-63.

Moleiro, C., Madureira, S., Verdelho, A., Ferro, J. M., Poggesi, A., Chabriat, H. et al. (2013). Confirmatory factor analysis of the Neuropsychological Assessment Battery of the LADIS study: A longitudinal analysis. *Journal of Clinical and Experimental Neuropsychology*, *35*(3), 269-278.

Montano, D. E., & Kasprzyk, D. (2015). Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health Behavior: Theory, Research and Practice* (pp. 95-120). San Fransisco: Jossey-Bass.

Moore, T. M., Reise, S. P., Gur, R. E., Hakonarson, H., & Gur, R. C. (2015). Psychometric properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology*, *29*(2), 235.

Moore, T. (2019). Description of emotional differentiation reliabilities.

Mota, N., Parada, M., Crego, A., Doallo, S., Caamaño-Isorna, F., Rodríguez Holguín, S. et al. (2013). Binge drinking trajectory and neuropsychological functioning among university students: a longitudinal study. *Drug Alcohol Dependence*, *133*(1), 108-114.

Murray, B. G. (2001). Are ecological and evolutionary theories scientific. *Biological Reviews*, *76*(2), 255-289.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*.

National Research Council. (1986). *Ecological knowledge and environmental problem-solving: concepts and case studies*. Washington, DC: National Academies Press.

Neuropsychology Section at University of Pennsylvania. (2012). *Penn Web-Based Computerized Neurocognitive Battery (WebCNP) Test Descriptions and Scoring Variables EVIONI2 Battery*. Pittsburgh, PA: University of Pennsylvania.

Newby, R. F., Hallenbeck, C. E., & Embretson, S. (1983). Confirmatory factor analysis of four general neuropsychological models with a modified Halstead-Reitan Battery. *Journal of Clinical and Experimental Neuropsychology*, *5*(2), 115-133.

Nguyen-Louie, T. T., Tracas, A., Squeglia, L. M., Matt, G. E., Eberson-Shumate, S., & Tapert, S. F. (2016). Learning and Memory in Adolescent Moderate, Binge, and Extreme-Binge Drinkers. *Alcoholism: Clinical and Experimental Research*, *40*(9), 1895-1904.

Nigg, C. R., Allegrante, J. P., & Ory, M. (2002). Theory-comparison and multiple-behavior research: common themes advancing health behavior research. *Health Education Research*, *17*, 670-679.

Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, *10*, 53.

Noar, S. M., Benac, C. N., & Harris, M. S. (2007). Does tailoring matter? meta-analytic review of tailored print health behavior change interventions. *Psychological Bulletin*, *133*(4), 673.

Noar, S. M., & Zimmerman, R. S. (2005). Health behavior theory and cumulative knowledge regarding health behaviors: are we moving in the right direction. *Health Education Research*, *20*(3), 275-290.

Norman, P., & Conner, M. (2006). The theory of planned behaviour and binge drinking: assessing the moderating role of past behaviour within the theory of planned behaviour. *British Journal of Health Psychology*, *11*(Pt 1), 55-70.

Norman, P. (2011). The theory of planned behavior and binge drinking among undergraduate students: assessing the impact of habit strength. *Addictive Behaviors*, *36*(5), 502-507.

Oechel, W. C., Hastings, S. J., Vourlrtis, G., Jenkins, M., Riechers, G., & Grulke, N. (1993). Recent change of arctic tundra ecosystems from a net carbon dioxide sink to a source. *Nature*, *361*(6412), 520.

Ogden, J. (2003). Some problems with social cognition models: A pragmatic and conceptual analysis. *Health Psychology*, *22*(4), 424.

Oscar-Berman, M., & Marinkovic, K. (2003). Alcoholism and the brain: an overview. *Alcohol Research and Health*, *27*(2), 125-133.

Oscar-Berman, M., Valmas, M. M., Sawyer, K. S., Ruiz, S. M., Luhar, R. B., & Gravitz, Z. R. (2014). Profiles of impaired, spared, and recovered neuropsychologic processes in alcoholism. In *Handbook of Clinical Neurology* (pp. 183-210). Elsevier.

Oster, G. (1981). Predicting populations. *American Zoologist*, *21*(4), 831-844.

Painter, J. E., Borba, C. P., Hynes, M., Mays, D., & Glanz, K. (2008). The use of theory in health behavior research from 2000 to 2005: a systematic review. *Ann Behav Med*, *35*(3), 358-362.

Parast, L., McCaffrey, D. F., Burgette, L. F., de la Guardia, F. H., Golinelli, D., Miles, J. N. V. et al. (2017). Optimizing variance-bias trade-off in the TWANG package for estimation of propensity scores. *Health Services and Outcomes Research Methodology*, *17*(3-4), 175-197.

Patt, V. M., Brown, G. G., Thomas, M. L., Roesch, S. C., Taylor, M. J., & Heaton, R. K. (2017). Factor analysis of an expanded Halstead-Reitan Battery and the structure of neurocognition. *Archives of Clinical Neuropsychology*, *33*(1), 79-101.

Paunesku, D., Walton, G. M., Romero, C., Smith, E. A., Yeager, G. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science*, *26*(6), 784-793.

Perry, G., & Pianka, E. R. (1997). Animal foraging: past, present, and future. *Trends in Ecology & Evolution*, *12*(9), 360-363.

Peskin, M. F., Hernandez, B. F., Gabay, E. K., Cuccaro, P., Li, D. H., Ratliff, E. et al. (2017). Using intervention mapping for program design and production of iCHAMPSS: an online decision support system to increase adoption, implementation, and maintenance of evidence-based sexual health programs. *Frontiers in Public Health*, *5*, 203.

Pfefferbaum, A., Kwon, D., Brumback, T., Thompson, W. K., Cummins, K., Tapert, S. F. et al. (2017). Altered brain developmental trajectories in adolescents after initiating drinking. *American Journal of Psychiatry*, appi. ajp. 2017.17040469.

Pickett, S. T. A., Kolasa, J., & Jones, C. G. (2010). *Ecological understanding: the nature of theory and the theory of nature*. Elsevier.

Pierce, G. J., & Ollason, J. G. (1987). Eight reasons why optimal foraging theory is a complete waste of time. *Oikos*, 111-118.

Poon, K. (2017). Hot and Cool Executive Functions in Adolescence: Development and Contributions to Important Developmental Outcomes. *Front Psychol*, *8*, 2311.

Popper, K. (2014). *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Routledge.

Popper, K. R. (1994). *The Myth of the Framework: in Defense of Science and Rationality*. London: Routledge.

Porter, S. R., & Pryor, J. (2007). The effects of heavy episodic alcohol use on student engagement, academic performance, and time use. *Journal of College Student Development*.

Pot, M., Ruiter, R. A. C., Paulussen, T. W. G. M., Heuvelink, A., de Melker, H. E., van Vliet, H. J. A. et al. (2018). Systematically developing a web-based tailored intervention promoting HPV-vaccination acceptability among mothers of invited girls using intervention mapping. *Frontiers in Public Health*, *6*.

Potochnik, A., Colombo, M., & Wright, C. (2018). *Recipes for Science: An Introduction to Scientific Methods and Reasoning*. Routledge.

Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological methods*, *17*(1), 1.

Prestwich, A., Sniehotta, F. F., Whittington, C., Dombrowski, S. U., Rogers, L., & Michie, S. (2014). Does theory influence the effectiveness of health behavior interventions? meta-analysis. *Health Psychology*, *33*(5), 465.

Project MATCH Research Group. (1998). Therapist effects in three treatments for alcohol problems. *Psychotherapy Research*, *8*(4), 455-474.

Pulido, C., Brown, S. A., Cummins, K., Paulus, M. P., & Tapert, S. F. (2010). Alcohol cue reactivity task development. *Addictive Behaviors*, *35*(2), 84-90.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Dev Rev*, *41*, 71-90.

Pyke, G. (2019). Optimal Foraging Theory: An Introduction. In *Encyclopedia of Animal Behavior, 2nd Edition, Volume 2* (pp. 111-117). Cambridge, MA: Academic Press.

Pyke, G. H. (1984). Optimal foraging theory: a critical review. *Annual Review of Ecology and Systematics*, *15*(1), 523-575.

Raftery, A. (1995). Baysian model selection in social research. *Social Methodology*, *25*, 111-163.

Reback, C. J., Fletcher, J. B., Shoptaw, S., & Mansergh, G. (2015). Exposure to theory-driven text messages is associated with HIV risk reduction among methamphetamine-using men who have sex with men. *AIDS and Behavior*, *19*(2), 130-141.

Rehm, J., Gmel, G. E., Gmel, G., Hasan, O. S., Imtiaz, S., Popova, S. et al. (2017). The relationship between different dimensions of alcohol use and the burden of disease-an update. *Addiction*.

Rehm, J., & Imtiaz, S. (2016). A narrative review of alcohol consumption as a risk factor for global burden of disease. *Subst Abuse Treat Prev Policy*, *11*(1), 37.

Rehm, J., Mathers, C., Popova, S., & Thavorncharoensap…, M. (2009). Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *The Lancet*.

Renshaw, E. (1993). *Modeling biological populations in space and time*. Cambridge, UK: Cambridge University Press.

Ridgeway, G., McCaffrey, D. F., Morral, A. R., Burgette, L. F., & Griffin, B. A. (2014). *Toolkit for weighting and analysis of nonequivalent groups: a tutorial for the R TWANG package*. Santa Monica, CA: RAND Corporation.

Romesburg, H. C. (1981). Wildlife science: gaining reliable knowledge. *The Journal of Wildlife Management*, 293-313.

Rosenbaum, P. R. (2009). Design of Observational Studies.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118-137.

Rothman, A. J. (2004). “Is there nothing more practical than a good theory?”: Why innovations and advances in health behavior change will arise if interventions are used to test and refine theory. *International Journal of Behavioral Nutrition and Physical Activity*, *1*, 11.

Rothman, A. J. (2011). Be prepared: capitalizing on opportunities to advance theory and practice. *Journal of Public Health Dentistry*, *71 Suppl 1*, S49-50.

Ryan, C., & Butters, N. (1980). Further evidence for a continuum-of-impairment encompassing male alcoholic Korsakoff patients and chronic alcoholic men. *Alcoholism: Clinical and Experimental Research*, *4*(2), 190-198.

Sagoff, M. (2016). Are there general causal forces in ecology. *Synthese*.

Salling, M. C., Jane Skelly, M., Avegno, E., Regan, S., Zeric, T., Nichols, E. et al. (2018). Alcohol consumption during adolescence in a mouse model of binge drinking alters the intrinsic excitability and function of the prefrontal cortex through a reduction in the hyperpolarization-activated cation current. *The Journal of Neuroscience*.

Sallis, J. F., Owen, N., & Fisher, E. (2015). Ecological models of health behavior. *Health Behavior: Theory, Research, and Practice*, *5*, 43-64.

Savi, A. O., Marsman, M., van der Maas, H. L. J., & Maris, G. K. J. (2019). The wiring of intelligence. *Perspectives on Psychological Science*, *14*(6), 1034-1061.

Schaefer, B. A., & McDermott, P. A. (1999). Learning behavior and intelligence as explanations for children's scholastic achievement. *Journal of School Psychology*, *37*(3), 299-313.

Scheiner, S. M. (2012). The multiple roles of theory: a reply to gorelick. *Ideas in Ecology and Evolution*, *5*.

Scheiner, S. M. (2013). The ecological literature, an idea-free distribution. *Ecology Letters*, *16*(12), 1421-1423.

Scheiner, S. M., & Willig, M. R. (2011a). A general theory of ecology. In *The Theory of Ecology*. Chicago: University of Chicago Press.

Scheiner, S. M., & Willig, M. R. (2011b). *The Theory of Ecology*. Chicago: University of Chicago Press.

Schmitt, T. A., Sass, D. A., Chappelle, W., & Thompson, W. (2018). Selecting the "best" factor structure and moving measurement validation forward: An illustration. *Journal of personality assessment*, *100*(4), 345-362.

Schnute, J. T., & Richards, L. J. (2001). Use and abuse of fishery models. *Canadian Journal of Fisheries and Aquatic Sciences*.

Schulenberg, J., Bachman, J. G., O'Malley, P. M., & Johnston, L. D. (1994). High school educational success and subsequent substance use: a panel analysis following adolescents into young adulthood. *J Health Soc Behav*, *35*(1), 45-62.

Schuler, M. S., Chu, W., & Coffman, D. (2016). Propensity score weighting for a continuous exposure with multilevel data. *Health Serv Outcomes Res Methodol*, *16*(4), 271-292.

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*.

Shrader-Frechette, K. (2008). *A companion to environmental philosophy*. John Wiley & Sons.

Singleton, R. A., & Wolfson, A. R. (2009). Alcohol consumption, sleep, and academic performance among college students. *Journal of Studies on Alcohol and Drugs*.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, *75*(3), 417-453.

Slobodkin, L. B. (1965). On the present incompleteness of mathematical ecology. *American Scientist*, *53*(3), 347-357.

Smedslund, G. (2000). A pragmatic basis for judging models and theories in health psychology: The axiomatic method. *Journal of Health Psychology*, *5*(2), 133-149.

Smit, F., Lokkerbol, J., Riper, H., Majo, C., Boon, B., & Blankers, M. (2011). Modeling the cost-effectiveness of health care systems for alcohol use disorders: how implementation of eHealth interventions improves cost-effectiveness. *Journal of Medical Internet Research*, *13*(3), e56.

Smith, F. E. (1952). Experimental methods in population dynamics: a critique. *Ecology*, *33*(4), 441-450.

Smith, F. E. (1976). Ecology: progress and self-criticism. *Science*.

Sneider, J. T., Cohen-Gilbert, J. E., Crowley, D. J., Paul, M. D., & Silveri, M. M. (2013). Differential effects of binge drinking on learning and memory in emerging adults. *Journal of Addiction Research & Therapy*.

Sniehotta, F. F., Presseau, J., & Araújo-Soares, V. (2014). Time to retire the theory of planned behaviour. *Health Psychology Review*, *8*(1), 1-7.

Spada, M. M., Albery, I. P., & Moss, A. C. (2015). Contemporary perspectives on cognition in addictive behaviors. *Addictive Behaviors*, *44*, 1-2.

Spear, L. P. (2018). Effects of adolescent alcohol consumption on the brain and behaviour. *Nat Rev Neurosci*, *19*(4), 197-214.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.

Squeglia, L. M., Pulido, C., Wetherill, R. R., Jacobus, J., Brown, G. G., & Tapert, S. F. (2012a). Brain response to working memory over three years of adolescence: influence of initiating heavy drinking. *Journal of studies on alcohol and drugs*, *73*(5), 749-760.

Squeglia, L. M., Rinker, D. A., Bartsch, H., Castro, N., Chung, Y., Dale, A. M. et al. (2014). Brain volume reductions in adolescent heavy drinkers. *Developmental cognitive neuroscience*, *9*, 117-125.

Squeglia, L. M., & Gray, K. M. (2016). Alcohol and drug use and the developing brain. *Current Psychiatry Reports*, *18*(5), 46.

Squeglia, L. M., Pulido, C., Wetherill, R. R., Jacobus, J., Brown, G. G., & Tapert, S. F. (2012b). Brain response to working memory over three years of adolescence: influence of initiating heavy drinking. *J Stud Alcohol Drugs*, *73*(5), 749-760.

Squeglia, L. M., Spadoni, A. D., Infante, M. A., Myers, M. G., & Tapert, S. F. (2009b). Initiating moderate to heavy alcohol use predicts changes in neuropsychological functioning for adolescent girls and boys. *Psychol ogy of Addictive Behaviors*, *23*(4), 715-722.

Squeglia, L. M., Spadoni, A. D., Infante, M. A., Myers, M. G., & Tapert, S. F. (2009b). Initiating moderate to heavy alcohol use predicts changes in neuropsychological functioning for adolescent girls and boys. *Psychol Addict Behav*, *23*(4), 715-722.

Staley, K. W. (2014). *An Introduction to the Philosophy of Science*. Cambridge, UK: Cambridge University Press.

StataCorp. (2017). Stata Statistical Software: Release 15.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of consumer research*, *25*(1), 78-90.

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, *19*(1), 80-90.

Sterman, J. D. (2006). Learning from evidence in a complex world. *American Journal of Public Health*, *96*(3), 505-514.

Strathdee, S. A., Hallett, T. B., Bobrova, N., Rhodes, T., Booth, R., Abdool, R. et al. (2010). HIV and risk environment for injecting drug users: the past, present, and future. *The Lancet*, *376*(9737), 268-284.

Sullivan, E. V., Deshmukh, A., Desmond, J. E., Lim, K. O., & Pfefferbaum, A. (2000). Cerebellar volume decline in normal aging, alcoholism, and Korsakoff's syndrome: Relation to ataxia. *Neuropsychology*, *14*(3), 341.

Sullivan, E. V., Brumback, T., Tapert, S. F., Fama, R., Prouty, D., Brown, S. A. et al. (2016). Cognitive, Emotion Control, and Motor Performance of Adolescents in the NCANDA Study: Contributions From Alcohol Consumption, Age, Sex, Ethnicity, and Family History of Addiction. *Neuropsychology*.

Sullivan, E. V., Brumback, T., Tapert, S. F., Prouty, D., Fama, R., Thompson, W. K. et al. (2017). Effects of prior testing lasting a full year in NCANDA adolescents: Contributions from age, sex, socioeconomic status, ethnicity, site, family history of alcohol or drug abuse, and baseline performance. *Dev Cogn Neurosci*, *24*, 72-83.

Tapert, S. F., & Brown, S. A. (1999). Neuropsychological correlates of adolescent substance abuse: four-year outcomes. *Journal of International Neuropsychological Society*, *5*(6), 481-493.

Tapert, S. F., Granholm, E., Leedy, N. G., & Brown, S. A. (2002). Substance use and withdrawal: neuropsychological functioning over 8 years in youth. *Journal of International Neuropsychological Society*, *8*(7), 873-883.

Taylor, N., Conner, M., & Lawton, R. (2012). The impact of theory on the effectiveness of worksite physical activity interventions: a meta-analysis and meta-regression. *Health Psychology Review*, *6*(1), 33-73.

Thompson, P. (1989). *The Structure of Biological Theories*. SUNY Press.

Thompson, W. K., Barch, D. M., Bjork, J. M., Gonzalez, R., Nagel, B. J., Nixon, S. J. et al. (2019). The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. *Developmental Cognitive Neuroscience*, *36*, 100606.

Tomlinson, K. L., Cummins, K. M., & Brown, S. A. (2013). Social anxiety and onset of drinking in early adolescence. *Journal of Child & Adolescent Substance Abuse*, *22*(2), 163-177.

Trafimow, D. (2015). Rational Actor Theories. In B. Gawronski & G. V. Bodenhausen (Eds.), *Theory and Explanation in Social Psychology* (pp. 245-267). New York: Guilford Publications.

Twenge, J. M., & Park, H. (2019). The Decline in Adult Activities Among U.S. Adolescents, 1976–2016. *Child Development*, *90*(2), 638-654.

van der Maas, H., Kan, K.-J., Marsman, M., & Stevenson, C. E. (2017). Network Models for Cognitive Development and Intelligence. *Journal of Intelligence*, *5*(2), 16.

Van Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*, *43*(2), 225-243.

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J. et al. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Annals of Internal Medicine*, *147*(8), 163-194.

Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. *Journal of Operations Management*, *16*(4), 361-385.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psycholical Science*, *7*(6), 632-638.

Wangersky, P. J. (1978). Lotka-Volterra population models. *Annual Review of Ecology and Systematics*, *9*(1), 189-218.

Watkins, M. W., Lei, P.-W., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, *35*(1), 59-68.

Weber, S. (2010). bacon: An effective way to detect outliers in multivariate data using Stata (and Mata). *The Stata Journal*, *10*(3), 331-338.

Wechsler, D. (2008). Wechsler Adult Intelligence Scale: Fourth Edition (WAIS-IV).

Weisberg, M. (2006). Forty years of 'the strategy': Levins on model building and idealization. *Biology and Philosophy*, *21*(5), 623-645.

Weisberg, M. (2007). Three kinds of idealization. *The journal of Philosophy*, *104*(12), 639-659.

Weissman, D. G., Schriber, R. A., Fassbender, C., Atherton, O., Krafft, C., Robins, R. W. et al. (2015). Earlier adolescent substance use onset predicts stronger connectivity between reward and cognitive control brain networks. *Dev Cogn Neurosci*, *16*, 121-129.

Werner, E. E., & Mittelbach, G. G. (1981). Optimal foraging: field tests of diet choice and habitat switching. *American Zoologist*, *21*(4), 813-829.

Whiteford, H. A., Degenhardt, L., Rehm, J., & Baxter…, A. J. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet*.

Windle, M., Spear, L. P., Fuligni, A. J., Angold, A., Brown, J. D., Pine, D. et al. (2008). Transitions Into Underage and Problem Drinking: Developmental Processes and Mechanisms Between 10 and 15 Years of Age. *Pediatrics*, *121*(Supplement 4), S273-S289.

Winther, R. G. (2018). The structure of scientific theories. *The Stanford Encyclopedia of Philosophy (Winter 2018 Edition)* Retrieved October 24, 2018, 2018, from https://plato.stanford.edu/entries/structure-scientific-theories/.

Winward, J. L., Bekman, N. M., Hanson, K. L., Lejuez, C. W., & Brown, S. A. (2014). Changes in emotional reactivity and distress tolerance among heavy drinking adolescents during sustained abstinence. *Alcoholism: Clinical and Experimental Research*, *38*(6), 1761-1769.

Winward, J. L., Hanson, K. L., Bekman, N. M., Tapert, S. F., & Brown, S. A. (2014). Adolescent heavy episodic drinking: neurocognitive functioning during early abstinence. *Journal of the International Neuropsychological Society*, *20*(2), 218-229.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

Worbe, Y., Irvine, M., Lange, I., Kundu, P., Howell, N. A., Harrison, N. A. et al. (2014). Neuronal correlates of risk-seeking attitudes to anticipated losses in binge drinkers. *Biol Psychiatry*, *76*(9), 717-724.

Wu, J. Y., Lin, J. J. H., Nian, M. W., & Hsiao, Y. C. (2017). A Solution to Modeling Multilevel Confirmatory Factor Analysis with Data Obtained from Complex Survey Sampling to Avoid Conflated Parameter Estimates. *Front Psychol*, *8*, 1464.

Ximénez, C. (2006). A Monte Carlo Study of Recovery of Weak Factor Loadings in Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(4), 587-614.

Yadav, S. K., Gupta, R. K., Garg, R. K., Venkatesh, V., Gupta, P. K., Singh, A. K. et al. (2017). Altered structural brain changes and neurocognitive performance in pediatric HIV. *NeuroImage: Clinical*, *14*, 316-322.

Yodzis, P. (1989). *Introduction to Theoretical Ecology*. San Fransisco: HarperCollins College Division.

Young, J. W. (1990). Are validity coefficients understated due to correctable defects in the GPA. *Research in Higher Education*, *31*(4), 319-325.

Zahr, N. M., & Pfefferbaum, A. (2017). Alcohol's effects on the brain: neuroimaging results in humans and animal models. *Alcohol Research*, *38*(2), 183-206.

Zhu, Y., Coffman, D. L., & Ghosh, D. (2015). A boosting algorithm for estimating generalized propensity scores with continuous treatments. *Journal of causal inference*, *3*(1), 25-40.

APPENDIX: MAJOR PHILOSOPHICAL VIEWS OF THEORY

Under the syntactic view, Theories are entities that can be reconstructed axiomatically (Winther, 2018). Theories are structured sentences that include a description of the relationships among conceptual entities, which are correspondence rules mapping the conceptual entities to things we can observe, and relationships among observable entities. The sentences may include the axioms, theorems, and laws associated with the Theory. Any difference in the formulation of the Theory renders it distinct from all others, because the Theory itself is the linguistic structure. Philosophers recognize that scientists may not present Theories with this surface structure. Theories only need to be susceptible to being projected into this form (Lambert & Brittan, 1992). Here, the structure of Theory is built on predicate logic. It follows that Theories are built with deductively linked sentences, with laws being derived from the conceptual sentences. Below laws and Theories, empirical generalizations are derivable from the abstract sentences. Finally, patterns in empirical observations follow from the empirical generalizations. In summary, conceptual relations are first stated, then rules for mapping the concepts to observable entities, and then relationships among observable entities follows from the conceptual and correspondence statements. It is the absolute distinction between conceptual and observable entities that is the principal challenge for the application of a purely syntactic view (Lambert & Brittan, 1992). An issue with this view is it can produce conceptual characterizations which may result in Theories differing only by linguistic choices applied during Theory construction (Thompson, 1989; Fraassen, 1989a). What we conceptualize in theoretical sentences and what we define as the corresponding observational entities within the Theory can challenge scientific progress. If scientific contexts change, how and what entities are measured may also change, which may result in Theories in the syntactic form no longer mapping to the research's observational targets. Syntactic structure has difficulty incorporating some types of conceptualizations. Complicated structures involving probability theory are not comfortably presented with the linguistic structure of syntactics (Suppes, 1957). Approaches based in mathematics became favored because of these philosophical limitations (Fraassen, 1989b; McEwan, ; Lambert & Brittan, 1992).

Under the semantic view, Theory is structured around a potential class of models built with mathematical logic (Thompson, 1989). Although they may coincide, these models can be distinguished from the sorts of models directly employed by scientists (Harré, 2004). We will denote the Theory model as $\mathfrak{M}$ (Harré, 2004). Theory consists of sentences with information about logical relationships of a general phenomenon, in abstract terms. $\mathfrak{M}$ provides information about the domain of discourse and extension of the predicates, indicating which instantiations of the predicates concur with the Theory. That information in $\mathfrak{M}$ is called a 'model' of the Theory if each of its sentences are true; To qualify as a Theory, its axioms must be specified such that $\mathfrak{M}$ is true. A Theory is then the set of consequences of the axioms expressed in a formal language, nominally mathematics, specifying $\mathfrak{M}$. $\mathfrak{M}$ helps provide meaning to a Theory. $\mathfrak{M}$ itself is of most immediate interest to logisticians and philosophers. It is generally tied to things of direct interest to scientists. $\mathfrak{M}$ can provide the constraints and structure of a particular model that might be used by researchers to make some prediction about a particular event (Boniolo, 2007). An introductory example can be found in Staley (Staley, 2014). It has been argued that the semantic approach has limitations, particularly at the point where the models are deemed representative of the real world. Neither the syntactic nor semantic view correspond to CDs that are a collection of antecedents of a phenomenon.

Using the semantic structure, Theory is considered to be a defined class of models that represent the target of scientific interest. These are the models, $\mathfrak{M}$, of Theory. The Theory also consists of models of experiment and models of data. The former outlines the study designs which fit with the Theory. Models of data are empirically estimated statistical models, which map onto some aspect of $\mathfrak{M}$. This is the Hierarchy of Models interpretation of semantically construed Theories. The Theory ties the logical models, which are susceptible to deductive evaluation and observations.