Title

Does Prosocial Identity Protect Justice-involved Youth from Reoffending? Testing Components of a Treatment-Relevant Theory

Author

Jian, Luyi

Publication Date

2024

Peer reviewed|Thesis/dissertation

Does Prosocial Identity Protect Justice-involved Youth from Reoffending? Testing Components of a Treatment-Relevant Theory

By

Luyi Jian

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Social Welfare
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Jennifer L. Skeem, Chair
Professor Susan Stone
Professor David J. Harding
Professor Stephen P. Hinshaw

Summer 2024

Abstract

Does Prosocial Identity Protect Justice-involved Youth from Reoffending?
Testing Components of a Treatment-Relevant Theory

by

Luyi Jian

Doctor of Philosophy in Social Welfare

University of California, Berkeley

Professor Jennifer L. Skeem, Chair

The Positive Youth Justice (PYJ) paradigm is gaining traction in the juvenile justice system. This paradigm shifts focus from youths' deficits to their strengths, opening new avenues to help young people build prosocial assets and desist from delinquent behavior. However, there is little conceptual or empirical guidance for translating the PYJ approach into real-world interventions. In this study, I focus on one promising construct within the Positive Youth Development framework—prosocial identity (i.e., the extent to which young people view themselves as prosocial)—and empirically test key hypotheses drawn from a conceptual model for intervention (Jian & Skeem, 2023). The central inquiries of the study include *how* prosocial identity should be defined as a treatment target in the juvenile justice context and under *what conditions* prosocial identity protects against reoffending. The primary aims are to first examine the basic protective effects of prosocial identity against reoffending in a justice-involved sample. Here, I include prosocial identity at the level of subcomponents—other-oriented (emphasizing benefits to others) and conventional (emphasizing law-abiding and conventional commitment), as well as holistically (prosocial identity as the combination of the two subcomponents). Second, I examine whether three factors—prosocial identity prominence (importance to the self), prosocial identity internal validation (confidence that the self can be achieved), and prosocial identity external validation (e.g., social support from peers and/or caregivers, school connectedness)—may moderate any protective effect of prosocial identity on recidivism. The secondary aim is to examine the generalizability of measures of prosocial identity that were developed with non-referred populations, to justice-involved youth.

I addressed these aims using data on a subsample of 760 court-referred youths. I conducted psychometric analyses and a series of progressively stringent analyses (from regression to causal modeling incorporating machine learning). The primary aims yielded three main findings. (1) Contrary to hypotheses, other-oriented identity, but not conventional identity, predicted recidivism independently. (2) There was no empirical support for the notion that prosocial identity *causally* protects against recidivism, after rigorously adjusting for covariates. (3) Among the proposed moderators, peer support (a form of external validation)

significantly moderated the relation between prosocial identity and recidivism (OR=0.99) in logistic regressions. The predicted probability of recidivism decreased as prosocial identity increased, only for those with moderate or high level of peer support. The moderating effect of peer support remained significant even after rigorously adjusting for all covariates, but not after using causal modeling. For the secondary aim, results indicate that prosocial identity measures developed with non-referred youth can be generalized to justice-involved youth.

These findings provide mixed support for hypotheses derived from the identity-based conceptual model. Research that directly tests aspects of the model is necessary to advance our understanding of whether and how deliberately *increasing* prosocial identity among justice-involved youth will reduce recidivism.

Acknowledgments

To the youth participants, staff, and research team who provided data support for this study. Without your joint efforts, this project could not have been completed as planned.

To my advisor and dissertation chair, Dr. Jennifer Skeem, thank you for your unreserved support and guidance throughout this journey. Your mentorship has profoundly impacted my personal and professional growth. I am grateful for your patience and belief in my abilities.

To the members of my dissertation committee, Dr. David Harding, Dr. Stephen Hinshaw, and Dr. Susan Stone, thank you for your constructive feedback, timely responses, and encouragement. Your expertise has been crucial in refining my thoughts and direction.

To the faculty, whose dedication to research and education has shaped me into a better scholar.

To my friends, for all the moments you listened to me and stood by me when I was lost.

To my family, you are my everything and the reason I continue to strive forward.

Statement

This dissertation is supported by the Administration for Children and Families (ACF) of the United States (U.S.) Department of Health and Human Services (HHS) as part of a financial assistance award (Grant #: 90PD0317) totaling $50,000 with 100 percent funded by ACF/HHS. The contents are those of the author and do not necessarily represent the official views of, nor an endorsement, by ACF/HHS, or the U.S. Government. For more information, please visit the ACF website, Administrative and National Policy Requirements.

Table of Contents

Introduction

Research and services for justice-involved youth are increasingly embracing Positive Youth Justice paradigms (PYJ; Butts et al., 2010; Haines & Case, 2018). Unlike traditional paradigms that emphasize reducing youth's deficits and risk factors for reoffending (e.g., Risk-Need-Responsivity model, Andrews, 2011), PYJ paradigms emphasize building on youth's strengths, developing their prosocial attributes and assets, and providing them with opportunities that can promote both desistance from antisocial behavior and overall healthy development (Butts, 2014; Butts et al., 2010; Haines & Case, 2018; Catalano et al., 2004).

Despite broad support for PYJ, direct application of these paradigms to juvenile justice interventions has been limited. It is challenging to translate PYJ into innovative, effective, and efficient interventions because the general paradigm is associated with a multitude of target constructs and components. According to Catalano et al (2004), "Positive Youth Development" programs target one or more of *fifteen* different constructs—including bonding, resilience, competency, self-determination, clear and positive identity, and prosocial norms and involvement. There is no empirical guidance on which of these constructs most strongly promotes desistance. In fact, there is not even consistent evidence that youths' strengths (or promotive factors) add value to their deficits (or risk factors), in predicting whether they reoffend (Barnes-Lee & Petkus, 2023). Although a leading PYJ paradigm targets only two constructs (learning/doing and attaching/belonging), it proposes that both diffuse assets be built across six different domains (i.e., work, education, relationships, community, health, creativity/art; Butts et al., 2010). As Butts (2014) acknowledged, this "requires justice professionals to look for those resources outside of their own agencies and budgets" (p. 2) and necessitates a system-wide approach that is "complicated to manage" (p. 4). In short, it is unclear what a PYJ intervention is—let alone, whether a PYJ intervention is effective.

To begin translating PYJ into innovative interventions that truly add value to juvenile justice, one approach is to unpack and investigate one particularly promising developmental construct and theoretical mechanism of change at a time. After all, "having too many goals and principles is akin to having none" (Research & Evaluation Center, 2023). In the current study, I begin to explore one promising construct within the positive youth development framework that could serve as a specific target of intervention: *prosocial identity*. A core purpose of adolescence is for a young person to establish their own identity and differentiate themselves from their parents or caregivers, as they gradually transform into an adult. As young people who are involved in the justice system cultivate their sense of self and place in the world, the process introduces both challenges *and* unique opportunities for intervention.

Prosocial identity can be defined as a domain-specific sense of self that focuses on the degree to which one is "prosocial" (Na & Paternoster, 2019; Paternoster & Bushway, 2008; Rocque et al., 2016). "Prosocial", in turn, can be conceptualized as a bipolar dimension that is anchored by antisocial (deviating from social norms and disregarding the rights of others) versus prosocial (supporting the social order and promoting others' welfare) tendencies (e.g., American Psychological Association, 2015; Eisenberg et al., 2015, p. 610; Pfattheicher et al., 2022). Theoretically, prosocial identity influences not only our internal processing, but also our

external behavior. According to leading psychological perspectives, identity motivates, directs, and regulates our behavior—so we behave in accordance with what would be socially expected for our salient identities (Stryker & Serpe, 1994), and prefer identity-congruent behavior over identity-incongruent behavior (Berkman et al., 2017a, 2017b; Oyserman, 2007). To the extent that a young person's dominant identity is prosocial, they will prefer conventional or prosocial behavior over antisocial behavior.

In keeping with this premise, compelling theories in criminology posit that shifts toward prosocial identity are a core mechanism of desistance from crime, among justice-involved adults (e.g., Paternoster & Bushway, 2008; Maruna, 2001). Shifts in identity and identity development are directly relevant to the period of adolescence, which typically begins with the onset of puberty (age 9 to 12) and ends when one begins taking on adult roles and responsibilities (age 18 to early 20's) (Crone & Dahl, 2012; Yeager et al., 2018). Nevertheless, the central role that prosocial identity could play in intervening with justice-involved youth has been underexplored.

Synthesizing theories and evidence from criminological-, developmental-, and clinical science, Jian & Skeem (2023) proposed a conceptual framework for identity-based intervention with at-risk and justice-involved youth. The central argument of the framework is that with individual effort and environmental support, a youth's identity can be shifted in the prosocial direction to promote desistance from antisocial behavior. The framework specifies three targets for change: the *content* of future possible selves (i.e., promoting hope for a future prosocial self, balanced by fear of a future antisocial self), prosocial identity *prominence* (or importance to the self), and prosocial identity *validation* (or confidence that the self can be achieved). The theoretical prediction is that the presence of prosocial identity—or the hoped-for prosocial self, featured by other-oriented content (i.e., involving an intention to benefit others, as emphasized by the developmental literature) and/or conventional content (i.e., being oriented toward traditional educational and vocational goals or family roles, as emphasized by the criminological literature)—will protect against antisocial behavior; and it will have particularly strong protective effects for youth with a high degree of prosocial identity prominence and validation and for youth with strong environmental support from family, peers, and school. In other words, this identity-based intervention framework hypothesized the theoretical mechanism of *how* prosocial identity works to prevent or reduce youth's antisocial behavior, and the *conditions under which* prosocial identity is protective.

This conceptual framework charts a pathway from the PYJ theory to intervention practice and provides empirically testable propositions. In the dissertation, I use data from a sample of court-referred youth (N = 760, ages 10 to 17) who were referred to community supervision in five counties in Pennsylvania between July 2021 and February 2023, to test key theoretical hypotheses of the framework. The dissertation represents an initial effort to empirically test the conceptual framework, with the aim of informing the development of identity-based interventions in the real world for at-risk and justice-involved youth.

In the rest of the Introduction[1], I review theories and evidence across disciplines that collectively suggest prosocial identity is a promising intervention target for juvenile justice programs. Then, I introduce the conceptual framework for intervention, adapted from Oyserman's (2007) work and that of others, to specify three targets for change: *content*, *prominence*, and *validation* of prosocial identity. This introduction will lay the groundwork for the motivation and research questions of the dissertation.

Prosocial Identity as a Promising Intervention Target

In this section, I integrate theories and evidence across different disciplines to offer three main reasons for starting with prosocial identity, as a particularly promising developmental construct that can inform juvenile justice programs. First, despite different operationalizations of the construct, emerging evidence across domains indicates that prosocial identity protects youth against antisocial behavior. Second, compelling theories in criminology prioritize the role of identity change as a core mechanism of desistance. Third, multidisciplinary theories and evidence suggest that identity is a developmentally "wise" intervention target for adolescents.

Evidence that Prosocial Identity Protects Youth Against Antisocial Behavior

Although research on the relation between prosocial identity and youth's antisocial behavior is relatively sparse, the results of several prospective studies suggest that prosocial identity protects against antisocial behavior—even though these studies involve different samples (unreferred youth, justice-involved youth, etc.) and various operationalizations of prosocial identity (single items, proxy measures, or validated scales). It is encouraging that the predictive utility of prosocial identity seems to generalize across a "heterogeneity of irrelevancies" like differences in measurement (Shadish, 1995), even though more research that uses validated measures of prosocial identity would be ideal.

Three groups of studies provide encouraging evidence. The first operationalized prosocial identity as either the reverse of the extent to which a youth viewed themself as a "troublemaker" and "delinquent" (Na & Paternoster, 2019), or the extent to which a youth viewed themself as being a conventionally "good person"—or law-abiding, trustful, honest, and not mean (Rocque et al., 2016). Using data from a large representative sample of 8th grade Korean youth who were surveyed every year for five years, Na and Paternoster (2019) examined the within-person association between changes in prosocial identity (see above) and changes in violent behavior, using the subsample of youth who had been involved in at least one violent incident (*N*=1,357, *M* age=13.8 at baseline). A one-unit increase in the prosocial identity scale (range=1-5) was related to a 19 percent decrease in the violence variety scale. Rocque and colleagues (2016) used data from a longer-term study that sampled adolescents from a birth cohort in New Jersey (*N*=447) and then followed them from age 12 to age 31, using five waves of surveys (three during adolescence; two during adulthood). Based on within-person analyses, the authors found that a one-unit increase in their prosocial identity scale

[1] Much content in the Introduction section is from a published paper by Jian & Skeem (2023).

(range=1-5) was associated with a significantly lower likelihood of engagement in crime (coefficient= -0.67).

A second group of studies used well-validated scales of moral identity to operationalize prosocial identity (Hardy et al., 2015; Kavussanu et al., 2015; Saulnier & Krettenauer, 2023). For example, Hardy et al. (2014) recruited a U.S. national sample of adolescents (*N*=384, ages 15-18) and their parents for an online study that examined the association between identity predictors (based on youth assessments) and behavioral outcomes (based on parent assessments). Youth completed the Moral Ideal Self Scale, which assesses the extent to which various moral traits (i.e., generous, respectful, truthful) describe the type of person they want to be; and a Moral Internalization Scale (Aquino & Reed, 2002), which captures how central a set of moral traits are, to their identity (e.g., *"Being someone with these characteristics is an important part of who I am."*). Parents rated their children on items drawn from a validated measure to assess delinquency. The authors found that both measures of moral identity significantly buffered against delinquency (*r*= -0.31, -0.33). Similar results were obtained with an independent sample of 510 youth, where youths' Moral Ideal Self ratings were moderately associated with parent ratings of aggression (Hardy et al., 2014).

Although the studies above focus on non-referred samples of youth, a third group of studies provide preliminary evidence that results generalize to justice-involved youth. First, based on a sample of 846 state-incarcerated youth (*M* age =17), Skeem and Kennealy (2015) used several items drawn from a risk assessment instrument to assess youths' prosocial identity and test its utility in protecting against serious or violent institutional infractions over a one-year period. The authors found that a one-point increase in the prosocial identity scale (range: -1 to 3) translated to 47% and 48% reduction in the rate of violent and serious infraction respectively. Moreover, prosocial identity significantly protected against institutional infractions even after adjusting for youth's scores on a comprehensive risk assessment instrument. Second, based on a large sample of justice-involved youth in Florida (*N*=222,640), Baglivio and colleagues (2014) found that prosocial attitudes (e.g., empathy for victims, respect for authority, favorable attitude toward law-abiding behavior, a sense of responsibility)—which is a concept closely related to prosocial identity—protected against subsequent reoffending over a one-year follow-up period.

In summary, despite variability in definitions and measurement of the construct, emerging evidence from several between-groups *and* within-participant (or repeated measures) studies indicates that prosocial identity relates inversely to antisocial behavior. Although most research has been conducted with non-referred youth, some studies suggest that prosocial identity may also protect against reoffending among youth who have been involved in violence and/or the justice system too.

Identity-Based Theories of Desistance from Criminal Behavior

In keeping with the empirical results summarized above, some criminological theories emphasize the role of prosocial identity in promoting desistance from crime. Although dominant theories of desistance heavily emphasize the role of external forces like social control (e.g., Laub & Sampson, 2003; Sampson & Laub, 1997), some alternative theories focus on shifts in identity as the primary—if not sufficient—cause of desistance (Bushway & Paternoster, 2013;

Giordano 2022; Giordano et al., 2002; LeBel et al., 2008; Maruna, 2001). According to these theories, exposure to prosocial environmental opportunities is critical—but there must be an intentional and *subjective* change in how one views oneself and crime, before one will actively seek and utilize those opportunities to transition to a noncriminal life.

Although it is not a theory of desistance *per se*, Oyserman and Markus (1990) offered a pioneering conceptualization of delinquency that features "possible selves" as a motivational source of young people's behavior. Possible selves are the future-oriented components of one's identity that include what one *expects* to become, and what one *fears* they might become. Based on a study of 238 youth who were sampled from juvenile justice and school settings, the authors found empirical support for a key hypothesis. According to that hypothesis, a given possible self protects a young person most strongly against delinquent behavior when it is *balanced* or offset by a countervailing possible self in the same domain (e.g., fearing becoming "a thief," but expecting to get along in school and get a job). When there is little balance between expected and feared selves, this decreases the motivational influence of the feared self on behavior. For example, if a young person cannot envision themselves as behaving differently in the future, they are unlikely to make choices that "prevent their feared selves from being realized" (p. 123).

Building on this conceptualization, Paternoster and Bushway (2008, 2013) offered an identity-based theory of desistance from crime that largely focuses on justice-involved adults. According to this theory, justice-involved adults have both *current* models of themselves as "criminal offenders, with a set of preferences and social networks consistent with that self" (p. 1103)—and *future* models of themselves that distill what they hope to become (the positive possible self) and what they worry about becoming (the feared possible self). A person's commitment to their current criminal self will weaken when they begin to perceive that the costs of commitment to that self outweighs its benefits. As their life's failures become salient and linked to identity, they believe they may in fact become the feared self—which provides motivation to change their preferences and social networks in the conventional direction. This, in turn, promotes desistance from criminal behavior.

This theory enjoys some empirical support. For example, Paternoster and colleagues (2016) used interview data from a sample that was predominantly composed of African American men who had been convicted of serious drug-related crimes (N=1,044, M age=29) and followed up for 20 years after release from prison. Nine months prior to, and six months after release, participants were asked whether they considered themselves "an addict". Based on their responses, participants were classified as viewing themselves as never an addict (endorsed neither time), a persistent addict (endorsed both times), a new addict (endorsed only post-release), or a reformed addict (endorsed only pre-release). Compared to "persistent addicts", those classified as "reformed addicts" had a significant longer time until re-arrest (HR=1.39, p<0.001). Similarly, those who sought any help for their drug problem had significantly longer survival times to re-arrest (HR=1.22, p<0.01) than those who did not. These findings suggest that participants whose identity changed (i.e., the "reformed addicts") and who took intentional actions in response to a favorable view of self (i.e., seeking help for substance use) were more likely to desist (Paternoster et al., 2016).

Paternoster and Bushway's identity-based theory is consistent with other perspectives that focus on the role of human agency in the desistance process. For example, Maruna (2001) intensively studied 65 people with chronic histories of offending. He found that those who continued their criminal careers often used a "condemnation narrative" to interpret their past and current criminal behavior, which tended to cast offending as their only choice. In contrast, those desisted from crime used a "redemption narrative" that cast their true self as a decent person and their past behavior as a special calling to make good in the future. This redemption narrative is a reflection that these people (who desisted from crime) emphasized the prosocial aspects of their identity and perceived themselves as being willing and able to "make good" in the future. In other words, the redemption narrative helped these people build a conventional or prosocial identity.

As suggested by this brief review, several theories articulate how identity can play a central role in promoting desistance from delinquent or criminal behavior. Nevertheless, these theories provide relatively little guidance on (a) whether identity-relevant processes can be deliberately changed and, if so, (b) how interventions could be mounted to achieve desirable justice outcomes.

Answers to these *whether* and *how* questions can be found in the psychological literature. There, rich theoretical narratives on identity-based motivation have been provided—including an identity hierarchy theory (Stryker, 1987), identity-based motivation model (Oyserman, 2007) and identity-value model (Berkman et al., 2017a, 2017b). These theories have been successfully applied to change youth's academic and health behaviors (Clark et al., 2005; Oyserman et al., 2002; Oyserman et al., 2006). Next, I turn to scholarly insights on how identity is a particularly promising intervention target for adolescents.

Scholarship Indicating that Identity Is a Promising Intervention Target for Youth

Although there are compelling identity-based theories of criminal desistance *and* evidence that prosocial identity protects youth against antisocial behavior, the promise of identity-based interventions for justice-involved youth has been underexplored. Largely outside the field of criminology, a body of literature suggests that identity is both a developmentally appropriate and psychologically wise target for intervening during adolescence.

First, prosocial identity is a *developmentally appropriate* intervention target for youth—in that it is both malleable and central to the needs and features of adolescence. Classic theories propose that adolescence is a key period of identity exploration and development (Erikson, 1968; Marcia,1980)—and developmental science generally supports such propositions (Crone & Dahl, 2012; Dahl et al., 2018; Pfeifer & Berkman, 2018; Pfeifer et al., 2007; Pfeifer et al., 2009). For example, Pfeifer and colleagues (2007, 2009) used functional neuroimaging to study developmental differences in how the brain responds to self-evaluative tasks, such as judging whether specific traits and attributes appropriately describe oneself. Compared to adults, children and adolescents responded more strongly to these tasks in brain structures associated with self-evaluative processes (the ventromedial prefrontal cortex and adjacent rostral/perigenual anterior cingulate cortex). Activities in these brain structures tend to increase from late childhood through middle adolescence and then either plateau or continue to increase (Pfeifer & Berkman, 2018, p.2), which suggests that adolescence is a period of

elevated self- and identity-related processing. More generally, reviews of developmental neuroscience indicate that adolescents are particularly sensitive to social and affective influences and broader changes in context—making adolescence a period of enhanced growth, learning, and adaptation during which "young lives can pivot rapidly in both negative and positive directions" (Crone & Dahl, 2012; Dahl et al., 2018, p.441). This evidence suggests that adolescence, particularly early-to-mid adolescence (see Odgers et al., 2022; Williams et al., 2019), offers a window of opportunity for intervening to promote prosocial development.

Second, prosocial identity is a *psychologically wise* intervention target for youth, in that it is a precise psychological target that is directly relevant to changing antisocial behavior (see above, "Evidence that Prosocial Identity Protects Youth Against Antisocial Behavior"). Based on a comprehensive review of relevant research, Walton and Wilson (2018) define wise interventions as approaches that target specific constructs to alter people's subjective meaning-making about themselves, others, and social situations—and change their behavior. Although these interventions tend to be brief or "light touch," they often achieve sustainable effects that are comparable to those of more prominent and complicated interventions (Walton & Wilson, 2018; Dweck & Yeager, 2019). Theoretically, the sustainable effects of wise interventions are generated from recursive cycles in which an initial change in subjective negative meanings (e.g., beliefs, interpretations, identities) alters a person's behavior in a manner that improves situations; and the improved situations, in turn, reinforce the change in subjective meanings that result in more adaptive behavior (Walton & Wilson, 2018).

In keeping with this conceptualization of wise interventions, theories of identity-based motivation posit that shifts in identity can have long-term impact on behavior. First, according to Oyserman et al.'s (2012) model, the effect of identity on behavior becomes stable and automatized when an individual repeatedly experiences psychologically similar situations that cue the same working identities and they take the same actions. This model explains why, in certain situations, people feel reflexively that their behavior is "right" (identity-consistent) or "wrong" (identity-inconsistent; Oyserman et al., 2012). Second, Pfeifer and Berkman (2018) propose an identity-value model that features a mutually reinforcing relation in which salient identities predict relevant behaviors, and identities that correspond to consistently chosen behaviors are also more valued by individuals over time. Thus, it is reasonable to assume that if a prosocial identity is successfully established and made dominant through intervention, it could protect youth against antisocial behavior in the long run—i.e., it could lead to desistance.

As suggested earlier, outside the domain of antisocial behavior, identity-based theories have been successfully applied to improve youth's academic and health outcomes. First, based on two samples of middle school students who were largely from minority groups (total *N*=472), Oyserman and colleagues (2002, 2006) developed and tested the School-to-Jobs program. This program consists of 9-14 group sessions that are designed to promote school-focused identity to improve students' academic performance. Students were randomly assigned to participate in the intervention group or a regular class, during their elective period. Intervention strategies are numerous and include the following: having students highlight their academic skills and abilities (to increase the salience of their school-focused identity); envision their desired possible selves (or adult identities); draw positive and negative forces (people and things) for attaining those desired selves; connect their school-focused identity to their desired

possible self through a timeline that includes nearer-term (one year) hoped-for and feared possible selves; identify strategies for attaining their school-focused identity and desired possible selves; and normalize and address difficulties they will encounter along the way while working to attain their school-focused identity (Oyserman et al., 2006). Results from the randomized controlled trial indicate that, over a one-year follow-up period, the program improved grades, increased school-relevant efforts, and—most relevant to the present review—reduced disruptive behavior at school (e.g., annoying peers, being verbally or physically abusive to the teacher), with standard effect sizes ranging from 0.15 to 0.36.

A second successful translation of identity-based theory to intervention focuses on youths' health outcomes. Based on a sample of Black middle-school youth (*N*=245), Clark et al (2005) developed and tested the Adult Identity Mentoring (AIM) intervention, which was designed to help youth from low-income families avoid premature sexual behavior. Across ten sessions that spanned a six-week period, the intervention focused on teaching youth skills to achieve positive future identities and involved no direct content on sexual behavior. Intervention strategies included elements like the School-to-Jobs program (e.g., envisioning desired possible selves and committing to one; drawing positive and negative forces for achieving the desired self; identifying skills, strengths, and strategies for actualizing one's future self, connecting the future self to current behavior through a timeline). However, unlike Schools-to-Jobs, AIM explicitly anchors the future possible self to an occupation (Clark et al., 2005). Compared to the control group that was offered standard health education classes, youth in the intervention group showed greater decreases in both their intention to have sex (16% versus 49%) and their self-reported sexual intercourse (decreased 10% versus -1%) over a 19-week period. Among male participants, the intervention's effect on abstinence from intercourse lasted over a one-year follow-up period.

The School-to-Jobs and AIM programs target academic- and health- behavior rather than antisocial behavior. Nevertheless, controlled trials indicate that both identity-based interventions improve young people's outcomes.

In sum, identity is a developmentally appropriate and psychologically wise target for intervention during adolescence. Successful identity-based interventions for youth have been developed outside the domain of antisocial behavior. The key question *is how* to deliberately intervene on relevant processes to foster youth's prosocial identity and promote desistance.

An Adapted Intervention Framework for Youth that Targets Prosocial Identity

Oyserman's identity-based motivation model (2007) and Berkman's identity value model (2017a, 2017b) propose different mechanisms through which identity influences behavior and identity-relevant behavior becomes stable over time. Nevertheless, two important points of agreement can be drawn from these models, along with relevant research and interventions (e.g., Bryan et al., 2011; Bryan et al., 2013; Oyserman et al., 2007). First, it is possible to change behavior by intentionally leveraging identity in the domain most relevant to that behavior. Second, changing identity can promote lasting change in behavior.

This convergence provides the foundation for an integrated conceptual framework that is specific to prosocial identity and youth's antisocial behavior. This framework is presented in

Figure 1. In the figure, the outer ellipse represents the social and cultural factors that encompass and shape identity construction and human behavior. The main part of the framework, which is in the middle of the ellipse from the left to right, illustrates the process by which a possible prosocial identity is initially motivated, gradually reinforced, and then works to promote prosocial behavior in situations where youth must make relevant choices.

The central hypothesis is that with individual effort and environmental support (shown at the bottom of the figure), a youth's identity can be shifted in the prosocial direction and doing so will promote desistance. The model specifies three intervention targets for establishing a coherent prosocial identity: *content* of the future possible self (promoting hope for a future prosocial self, balanced by fear of a future antisocial self), prosocial identity *prominence* (promoting its importance to the self), and prosocial identity *validation* (instilling confidence that the self can be achieved). In the remainder of this section, I describe the identity change process and then unpack each of the three targets.

Process of Initiating, Establishing and Reinforcing a Prosocial Identity

As shown in Figure 1, the initial motivation for leaning toward a possible prosocial identity is often provided by a *feared identity* or an antisocial self that the youth fears becoming (e.g., *"I will drop out of school and wind up in prison"*). A hoped-for self (e.g., *"I will graduate from high school and find a good job"*) will be a stronger incentive for desistance from delinquency when it is accompanied by a feared self (Oyserman and Markus, 1990). According to Bushway and Paternoster (2013), "at least initially, movement out of a deviant or a spoiled identity is more likely to be based on a motivation to avoid a feared self than on a desire to achieve a positive self" (p. 222), and a "balance" between negative and positive possible selves motivates and guides change.

Although a feared identity may tilt the balance toward a possible prosocial identity (through avoidance motivation), the prosocial identity will need to be reinforced over time to become established and achieve stable effects on behavior (through approach motivation). When a prosocial identity becomes dominant or overtakes a previously antisocial identity, identity-congruent behavior becomes more intuitive or natural—i.e., more driven by the mindset associated with prosocial identity (see Oyserman & Destin, 2010) and less dependent on deliberate evaluation that involves weighing the subjective values of different choice attributes such as rewards, costs and self-relevance (see Berkman et al., 2017a).

Three elements of prosocial identity—content, prominence, and validation—work together to establish and reinforce prosocial identity as a stable motivator for prosocial behavior (see Figure 1). As a youth becomes motivated to avoid a feared self and achieve a prosocial self, the prosocial identity may come to mind in a situation where they need to make a relevant choice (e.g., "*my friends offer me drugs after school*"). The content of prosocial identity—what prosocial generally means to *me* and what *my* possible prosocial selves are—is shaped by the youth's social environment, life experiences, and individual reflections on the self (Abrams & Aguilar, 2005; Clinkinbeard & Murray, 2011; Oyserman & Destin, 2010). But the specific meaning of prosocial identity at a given moment is also shaped by the situation (Oyserman, 2007; e.g., "*a 'good' person does not break school rules or laws*"). If a prosocial identity is prominent and validated, namely, if it feels more important than an antisocial

identity (e.g., "*I value being a good person more than being a drug user*") and also feels achievable (e.g., "*I can do it—I know how to turn down the offer without offending my friends*"), the youth is likely to act in accordance with the prosocial identity and associated mindset (e.g., "*I turn down the offer*").

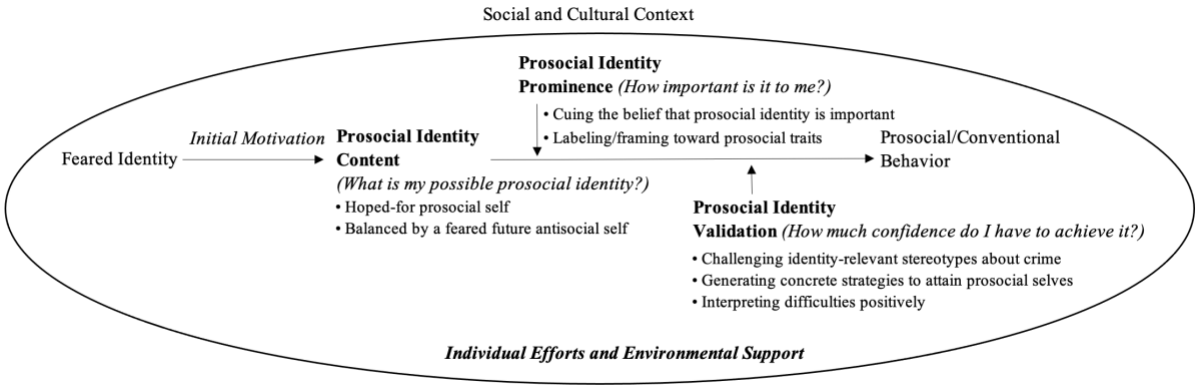

Figure 1. An Adapted Intervention Framework for Youth That Targets Prosocial Identity

Three Elements of Prosocial Identity that Can Be Leveraged by Interventions

Prosocial Identity Content. Identity content refers to the topics or issues that people consider when they think about who they are (McLean et al., 2019). Identity content is often studied by domain (e.g., school, work, family, politics; McLean et al., 2019). For a domain-specific identity (e.g., gender identity), identity content refers to the meaning people attribute to that identity (e.g., what it means to be a girl/boy; Burke, 2020; Burke & Tully, 1977). Although identities can be focused on the past (what used to be true of one), the present (what is true of one now), or the future (the person one wishes or fears to become) (Oyserman et al., 2012), the relevant target for intervention is the future-focused identity because possible identities motivate behavior change (Oyserman et al., 2006).

Thus, prosocial identity can be defined as a youth's conceptualization of the type of prosocial person they hope to become. Again, here prosocial is conceptualized as a bipolar dimension that is anchored by antisocial tendencies at one end (deviating from social norms and disregarding the rights of others) and prosocial tendencies at the other (supporting the social order and promoting others' welfare). Accordingly, the content of prosocial identity can vary widely, ranging from a possible self who achieves basic educational and vocational goals and takes on conventional adult roles and responsibilities (i.e., approximately in the "middle" of the prosocial spectrum), to a possible self who works hard to benefit others or is caring, contributing, helpful, and kind. This understanding of the content of prosocial identity is informed by how prosocial identity has been measured and studied differently across disciplines. In psychological research, validated measures have often been used to operationalize prosocial identity as moral identity, or the "extent to which people identify with, and are invested in, being a moral person and doing what is moral" or right (Hardy et al., 2014, p. 45). At the upper end, this is an other-oriented identity that includes an intention and reflects in behaviors to benefit others (Eisenberg et al., 2015; Kanacri et al., 2013). But in the criminological literature, ad-hoc proxy measures have typically been created to operationalize

prosocial identity as the extent to which people identify with, and are invested in, being an upstanding person and following laws and social norms (Bushway & Paternoster, 2013; Na & Paternoster, 2019; Rocque et al., 2016). Emphasis is placed on conventional behavior that shows a commitment to education, employment, and family goals. For example, Bushway and Paternoster (2013) describe a prototypic possible prosocial self as "working in a job (though perhaps for minimum wage), legitimately buying things for my family, owning a used car, and ceasing my life of drug use and crime" (p. 222).

Even when young people are not prompted to think about prosocial identity specifically, they often identify relevant possible selves. For example, based on a large sample of detained youth (*N*=543), Clinkinbeard and Murray (2011) found that youths' most common desired selves were relevant to school- (e.g., "*getting my GED*"), lifestyle- (e.g., a change of living situation), and vocational goals (e.g., "*working full time*"). An important question for identity-based interventions is whether practitioners should seek to establish hoped-for prosocial selves that are merely conventional (as emphasized in the criminological literature)—or should aim beyond that to establish moral, or explicitly other-oriented, values and tendencies (as emphasized in the psychological literature).

My hypothesis is that interventions are expected to promote desistance if they merely achieve conventional hoped-for selves and behavior. However, the level of a possible self's prosociality (or change in prosociality) that is necessary to promote desistance is an empirical question. In the dissertation, I attempt to answer this question by comparing measures of other-oriented identity and conventional identity constructed from the Moral Ideal Self (Hardy et al., 2014) scale, a well-validated self-report instrument for youth that assesses the extent to which 20 moral traits (i.e., generous, respectful, truthful) describe the type of person the participant wants to be. I examine which construct—other-oriented or conventional identity—is more relevant in protecting justice-involved youth from reoffending. Answers can help shape the interventions that are designed to promote prosocial identity for justice-involved youth.

Prosocial Identity Prominence. Individuals have multiple identities, some of which feel more important than others (McCall & Simmons 1978; Stryker & Serpe, 1994). Identity prominence is one's subjective sense of the importance or rank of an identity, in one's own identity hierarchy (Brenner et al., 2014; McCall & Simmons 1978; Stryker & Serpe, 1994). Prominence is related to "salience," or the likelihood that a given identity will be activated across social situations (Brenner et al., 2014; Burke, 2020; Oyserman & Destin, 2010). Prominence and salience are sometimes used as interchangeable terms—and measures of both constructs predict one's behavioral commitment to an identity (Stryker & Serpe, 1994). Given recent findings that suggest a causal ordering from prominence to salience (Brenner et al., 2014), I use "prominence" here.

Prominence can be conceptualized as the extent to which a prosocial identity feels important to *me*. I hypothesize that the effect of prosocial identity on behavior is moderated by its prominence, in situations that involve relevant choices (e.g., moral decision-making). This hypothesis is consistent with theories and research indicating that an identity's prominence modifies its impact on behavior (Burke, 2020; Cohn et al., 2015; Maitner et al., 2010; Marin et al., 2009; Oyserman et al., 2007; Stryker & Serpe, 1994). As the prominence of a given identity

increases, individuals are more likely to invest time acting on the identity-relevant role (e.g., Stryker & Serpe, 1994) and to endorse and apply identity-consistent attitudes and behavior (e.g., Oyserman et al., 2007). For example, in an experiment conducted with 182 adult male inmates, Cohn and colleagues (2015) used survey questions to manipulate the prominence of criminal identity by reminding the experimental group about their criminal behavior (e.g., "*What were you convicted of?*") while asking the control group questions that were unrelated to crime (e.g., "*How many hours per week do you watch television?*"). Participants then performed a task that provided an opportunity to cheat (i.e., misreport information) to earn money. Compared to the control group, the frequency of cheating was 60% higher in the group whose criminal identity had been made more prominent (Cohn et al., 2015). Still, the hypothesis that the effect of prosocial identity on behavior is moderated by its prominence remains to be tested among justice-involved youth.

It is worth noting that prosocial identity prominence can be correlated with prosocial identity content, but they are considered separate constructs because the literature suggests that they develop with age and associate with behaviors differently. Research shows that youth's moral ideal self—the type of moral person with a set of prosocial and conventional traits they want to be—does not correlate with age across adolescence (Hardy et al., 2014; Krettenauer, 2011; Pratt et al., 2003). Yet moral internalization—the degree to which becoming a moral ideal self is an important part of one's personal identity—increases with age (Hardy et al., 2014). Moral ideal self and moral internalization were positively correlated in the study (r = 0.48 to 0.51), but the two constructs are associated with different behavioral outcomes in early and late adolescence. For early adolescents, moral ideal self predicted aggression and altruism, while moral internalization predicted none of these outcomes. For late adolescents, moral ideal self predicted altruism, while moral internalization predicted aggression (Hardy et al., 2014). Thus, the content and prominence of prosocial identity may develop and link to behavior differently in adolescence.

Prosocial Identity Validation. Validation of prosocial identity refers to the extent to which youth believe they can achieve a hoped-for prosocial identity—i.e., the extent to which the hoped-for identity feels attainable and the matching feared self feels avoidable (see Oyserman & Markus, 1990). Given past theory and research (e.g., Abrams & Aguilar, 2005; Cuevas et al., 2017; Goodson & Morash, 2017; Oyserman & Markus, 1990; Oyserman & Saltz, 1993), I hypothesize that prosocial identity must be validated to have a meaningful and sustainable effect on behavior. There is indirect empirical support for this hypothesis. Based on a sample of 12,955 justice-involved youth, Cuevas and colleagues (2017) measured young people's sense of *prosocial self-efficacy*—i.e., their belief in their ability to live a prosocial life and to control antisocial behavior. The authors found that prosocial self-efficacy predicted youths' earlier release from residential placement and protected them against subsequent delinquent behavior (OR = 0.91, CI = [0.85–0.97], $p < .05$).

The validation process requires not only individual effort (i.e., *internal* validation, including self-perceptions of one's ability to achieve a prosocial identity) but also environmental support (i.e., *external* validation, including opportunities for achieving a prosocial identity and supportive messages from parents, teachers, and significant others). Inclusion of external validation is consistent with a basic premise of wise interventions—i.e., altering someone's

subjective meanings "will improve outcomes only when other aspects of the system necessary for improvement are in place" (Walton & Wilson, 2018, p. 620). Prosocial identity validation (plus prominence) is important for interventions because they can be shaped by the environment and deliberate efforts to increase them.

In summary, the theoretical propositions of the conceptual framework for an identity-based intervention are that prosocial identity content, prominence, and validation work together to motivate and sustain youth's prosocial and conventional behavior, and in the same vein reduce youth's antisocial behavior. Prosocial identity content represents the presence of prosocial identity, which can be featured by prosociality and/or conventionality. Prosocial identity prominence and validation represent the process of establishing a coherent prosocial identity and function potentially as moderators that influence the strength of protection that prosocial identity has on antisocial behavior. With empirical support, we may have more confidence in developing an identity-based intervention that focuses on fostering the presence of a possible prosocial identity and increasing its degree of prominence and validation, to reduce youth's antisocial behavior.

Research Aims and Hypotheses

The dissertation uses secondary data to examine some initial hypotheses derived from the conceptual model. Rather than conducting a rigorous test of the model with specifically designed measures for the construct of interest, I use both validated and proxy measures for key constructs based on the available data. Also, rather than attempting to manipulate aspects of prosocial identity, I use observational data to explore the utility of prosocial identity and other constructs measured at a single time point, in predicting reoffending. Given the proxy nature of some of the measures and the absence of an experimental design testing intervention strategies relevant to prosocial identity, the current study is exploratory in nature. It aims to understand relations among the constructs in the conceptual model.

The **primary aims** are to (a) examine the basic protective effect of prosocial identity, as well as other-oriented versus conventional identity, in a justice-involved sample, and (b) explore whether three factors—prominence, internal validation, and external validation of prosocial identity—moderate any protective effect of prosocial identity on delinquent behavior (e.g., recidivism). A specific statement of each question and hypothesis is provided below.

- **Question 1: How should prosocial identity be defined as a treatment target for juvenile justice interventions?** Specifically, compared to identity content that is conventional (i.e., emphasizes law-abiding and conventional commitment), does identity content that is other-oriented (i.e., emphasizes benefiting others) add value in protecting justice-involved youth against reoffending?
    - Hypothesis: Other-oriented identity will add no significant incremental utility to conventional identity in protecting justice-involved youth from reoffending.
    - Implications: If the hypothesis holds, justice interventions aimed at fostering a prosocial identity should start with assisting youth develop possible future self-images that are conventional—arguably a more reachable goal.

- **Question 2: Does the protective effect of prosocial identity depend on youths' sense of prominence of prosocial identity?**
    - Hypothesis: Prosocial identity will protect against reoffending but will have particularly strong effects among youth with a high degree of internalized prosocial identity.
    - Implications: If the hypothesis holds, justice interventions aimed at fostering a prosocial identity should not only prime youth with positive self-images but also emphasize the personal significance of possessing such prosocial traits. For potential intervention strategies, see Jian & Skeem (2023).

- **Question 3: Does the protective effect of prosocial identity depend on youths' sense of self-efficacy (as a proxy for prosocial identity validation)?**
    - Hypothesis: Prosocial identity will protect against reoffending but with particularly strong effects in youth with a high degree of general self-efficacy.
    - Implications: If the hypothesis holds, justice interventions aimed at fostering a prosocial identity should also help youth believe that a prosocial identity is achievable. This could include addressing stereotypes related to crime that affect their identity, devising tangible strategies for attaining a prosocial identity, and positively framing challenges encountered along the path.

- **Question 4: Does the protective effect of prosocial identity depend on youths' degree of social support and school connectedness?**
    - Hypothesis: The protective effect of a prosocial identity will be especially pronounced in youth with more social support and school connectedness, as opposed to those with less environmental support.
    - Implications: If the hypothesis holds, it would provide evidence for the synergistic effects of prosocial identity and environmental support on behavioral change (e.g., Paternoster & Bushway, 2008). Accordingly, justice interventions focusing on prosocial identity should incorporate efforts to provide environmental support, such as increasing social support and enhancing school connectedness, to more effectively shield youth from engaging in future delinquent behavior. In other words, programs can blend identity-focused and environmental-focused approaches to foster better justice outcomes.

The secondary aim of this dissertation is to examine the generalizability of measures of prosocial identity that were developed with non-referred populations, to justice-involved youth. This study appears to be the first application of these measures to justice-involved youth, which provides an opportunity to test the appropriateness of their use with this population. Addressing this secondary aim could pave the way for more valid assessments of identity among justice-involved youth.

In the next section, I explain how the study aims were addressed. I begin with a description of the data sources for this secondary analysis.

Method

Study Design and Data Source

The study uses an observational design. Secondary data are drawn from an ongoing project supported by the National Institute of Justice (NIJ). In this NIJ project, investigators from the University of Massachusetts Medical School (UMMS) and the School of Social Welfare at University of California, Berkeley (UCB) have partnered with juvenile justice agencies in three states—Pennsylvania (PA), Virginia (VA), and Wisconsin (WI)—to collect longitudinal data and conduct analyses that can inform the optimization of supervision and service strategies for justice-involved youth under community supervision. The UMMS research team has obtained full approval from their Institutional Review Board (IRB) and handles raw data with personal identifiers. The UCB research team receives and analyzes the de-identified data to answer research questions of the NIJ project. These de-identified data include the data used in the dissertation.

To obtain permission from the IRB at UCB for data use, the UCB research team submitted a protocol under the "Soc-Behav-Ed Exempt" category to the Committee for the Protection of Human Subjects (CPHS) at UCB in March 2023. The protocol was approved in April 2023.

As part of the broader NIJ project, a Protective Factor (PF) Survey was administered to youth referred to community supervision and who underwent a risk-need assessment per state policy from January 2020 to March 2023 in eleven counties across the three states. The PF survey included nine validated scales that measured youth's moral ideal self, moral internalization, life purpose, general self-efficacy, self-control, peer and caregiver support, school connectedness, social responsibility, and prosocial engagement in school and community. Youth completed the survey online around the intake of their community supervision. If they needed help reading, the survey was completed with assistance from staff who received training to ensure data collection was applied appropriately. Youth's ages at intake varied from 9 to 20 years old. Official recidivism records from a minimum 1-year follow-up were collected from the juvenile and criminal justice departments of the three states, respectively.

For the dissertation, I mainly used data from the baseline PF survey, baseline risk assessment, and 1-year recidivism for youth in five counties in PA who were enrolled in the NIJ project. The primary reason for using the PA data is that it is currently available for the completion of my dissertation, and its sample size meets the requirements for statistical power (see the next section, "Samples"). When all the baseline and recidivism data from the three states become available later this year, the analyses conducted in the dissertation will be replicated before the dissertation is transformed into publishable papers.

Samples

Multi-state Sample for Secondary Aim (Psychometric Analyses)

As introduced earlier, both validated (i.e., moral ideal self, moral internalization) and proxy measures (i.e., self-efficacy) of prosocial identity constructs were used in the study (see

next section, “Measures”). Before tackling the four research questions, psychometric analyses for these measures were conducted among a larger sample of 1,913 youths from all three states (PA, VA, WI) who completed the baseline PF survey and risk-need assessment. The characteristics of the larger sample (see Table 1 in Appendix B) aligned closely with those of the PA subsample used for addressing the primary aims.

The rationale for conducting psychometric analyses on this larger sample includes: 1) a larger sample size yields more accurate estimates of the measures’ reliability (e.g., internal consistency) and validity (e.g., construct validity); and 2) the baseline data for the larger sample was available one year earlier than the PA subsample—the latter could only be finalized after the recidivism follow-up data were available. For practical reasons, performing psychometric analyses on the larger sample aligns better with the timeline for completing the dissertation.

Pennsylvania Subsample for Primary Aims

The subsample of the PA data was selected for this study based on consideration of tradeoffs between data accuracy and sufficient statistical power. In the five counties in PA, a total of 1,293 youths completed the PF survey from July 2021 to February 2023 at baseline, spanning roughly a 1.5-year period. However, not all data from these youths could be used. To be included in the study, the youth needed to meet three criteria:

1) Being 17 or younger on the start date from when recidivism records were tracked. This is because recidivism records from the criminal justice system—for those who became older than 18 during the one-year follow-up period—were not currently available. Therefore, an age cutoff of 17 at baseline ensures that the one-year recidivism data were fully captured by the juvenile system. This removed 348 cases where youths’ ages were above 17 at baseline.
2) Not being censored during the study period. Five youths who were deceased, moved out of the state, or lost follow-up were removed from the analysis.
3) Having valid responses to the PF survey. Seven attention-check items were woven into the PF survey and served as the validity check for youths’ responses. An example of such items can be “*Are you willing to carefully read the instructions and answer all of the questions to the best of your ability?*” Responses are “Yes” or “No.” If a youth answered “No,” the item was marked as incorrect. The greater number of incorrect responses to these items may indicate that the youth did not complete the PF survey attentively, therefore their responses were likely unreliable. Among the 1,293 youths who completed the PF survey, 809 youths “passed” all seven attention-check items.

Using criteria 1) and 2), and relaxing the criterion 3) to include those who “passed” all *or* had only one incorrect response to the seven attention-check items in the PF survey yielded a sample of 760 youths, with the observed one-year recidivism rate of 13.8%. The corresponding minimal sample size with 80% statistical power at a significance level of 0.05 is 532 cases. Given that having only one incorrect response to attention-check items suggests that youth were

paying attention to the PF survey most of the time, I decided to use data from these 760 youths as the final sample.[2]

Table 1 below presents characteristics of the PA subsample. Compared to the larger sample in which psychometric analyses were conducted, the PA subsample is younger and has fewer high-risk youth (8.6% versus 12.4%).

Table 1. Characteristics of the PA subsample

	Mean (SD)	Range	*N*=760	
			n (valid %)	Missing
Age at Intake	15.1 (1.4)	10.2-17.0		
Early adolescence (<=14)			179 (23.6%)	
Middle adolescence (>14 & <=17)			581 (76.4%)	
Gender				
Male			533 (70.1%)	
Female			227 (29.9%)	
Race				
White			276 (36.3%)	
Latinx			102 (13.4%)	
Black			350 (46.1%)	
Other/Unsure			32 (4.2%)	
SDI score	51.4 (31.0)	1–100		1 (0.1%)
Low disadvantage [0, 33]			260 (34.3%)	
Moderate disadvantage [34, 66]			204 (26.9%)	
High disadvantage [67, 100]			295 (38.8%)	
Index/Current Referral (Primary)*				
Person			478 (62.9%)	
Property			94 (12.4%)	
Drug			120 (15.8%)	
Minor/Other			68 (8.9%)	
Index Referral was Violent			362 (47.6%)	

[2] Using criteria 1) and 2) and including those who "passed" all seven attention-check items resulted in a sample of 571 youths aged 10 to 17, who had the most reliable PF survey responses and one-year recidivism records. However, this sample size does not have sufficient statistical power to detect a small to medium relation—a point-biserial correlation of 0.18 or 0.24 based on findings from the literature (Brown et al., 2020; Hardy et al., 2015)—between protective measures and antisocial behavior or recidivism at a significance level of 0.05 and 80% power, after adjusting for the observed one-year recidivism rate of 12.2% among these 571 youths (measured by any petition, see the "Measures" section). The minimal sample size required for detecting a point-biserial correlation of 0.18 with a 12.2% recidivism rate is 590 cases (details for power analysis can be provided upon request).

Prior Referral				
Age at first offense	14.6 (1.5)	10.0–17.0		
Number of prior offenses	0.2 (0.7)	0–8		
Any violent prior referral			43 (5.7%)	
Risk Score	11.0 (6.6)	0–32		42 (5.5%)
Risk Level				42 (5.5%)
Low			335 (46.7%)	
Moderate			321 (44.7%)	
High			62 (8.6%)	

Measures

Prosocial Identity

As discussed earlier, one's future identity provides motivation to change current behavior that is incongruent with that identity, thus is the target of an identity-based intervention. The operationalization of prosocial identity in the study, therefore, focuses on its future-oriented facet.

The prosocial identity measures in the study do not include feared possible selves, which are an essential component of the conceptual model (see Figure 1 in the "Introduction" section). This specific, theory-tailored measure is unavailable in the secondary data used. Thus, the study can only assess the basic predictive utility of a hoped-for prosocial identity, not its potentially enhanced effects on delinquency desistance when balanced by a feared self.

Prosocial Identity Content. Prosocial identity was assessed with the Moral Ideal Self Scale (MISS, Hardy et al., 2014). The scale measures the extent to which 20 moral traits (*e.g., respectful, truthful, caring, generous*) describe the type of person a youth wants to be (response scale = 1/not at all to 7 /very much). The scale has been validated among adolescents in the general population (*N*=893, Hardy et al., 2014). In that sample, internal consistency for both total scores and subscale scores were high ($\alpha = 0.96$ and 0.92) and confirmatory factor analysis indicated that the proposed five-factor structure fit the data well (CFI = 0.99). Those five factors or subscales are:

- Loving/Caring (9 items): *generous, understanding, thankful, compassionate, loving, forgiving, considerate, caring, helpful*
- Honest/True (3 items): *truthful, true, loyal*
- Integrity (2 items): *stands up for his/her beliefs, follows values*
- Virtuous (2 items): *respectful, responsible*
- Knows/Choose Right (4 items): *good example, makes good choices, does good actions, has good values*

In the current study of justice-involved youth (multi-state sample, *N*=1,913), the MISS also demonstrated strong internal consistency for total scores ($\alpha = 0.94$). I used two subscales in the current study (rather than five subscales) because the first study aim required distinguishing between "other oriented" and "conventional" identity. To ensure this approach

was appropriate, I tested the fit of a two-factor model that divided the scales into "other oriented" (original Loving/Caring subscale) and "conventional" (original Honest/True, Integrity, Virtuous, and Knows/Choose Right subscales) identity. Results indicate that this two-factor model (CFI=0.92, RMSEA=0.07) fit the data as well as the original five-factor model (CFI=0.93, RMSEA=0.07). Given those results, I used each youth's average or sum of scores[3] on the 9-item Loving/Caring parcel as a measure of their other-oriented identity, and the average or sum of scores on other four parcels as a measure of their conventional identity to address the first study question.

For the remaining study questions (two-four), I used MISS total scores to operationalize prosocial identity holistically (including both other-oriented and conventional features). Not surprisingly, the correlation between the "other-oriented" and "conventional" factors are moderate to strong (τ=0.58, see Table 3 in the "Results" section). This suggests an overlap between the two factors. Depending on analytical approaches, this holistic prosocial identity can be a continuous variable based on the scores, or a dichotomous variable with two levels (low/high) based on 50% percentile (i.e., median).

Details on the factor structure and validity of the MISS with this sample of justice-involved youth are provide in Appendix B. As shown there, the MISS distinguishes between known groups (e.g., justice-involved youth obtained lower prosocial identity scores than non-referred samples) and manifests a pattern of divergent and convergent associations with other measures that are theoretically coherent.

Prosocial Identity Prominence. Prosocial identity prominence was measured with the 5-item Moral Internalization Scale. This scale assesses the extent to which being someone with a certain set of moral traits (i.e., *caring, compassionate, fair, friendly, generous, helpful, hardworking, honest*)—is central to one's personal identity (Aquino & Reed, 2002). Sample items in the scale include *"It would make me feel good to be a person who has these traits", "Having these traits is not really important to me",* and *"I strongly desire or want to have these traits"* (response scale = 1/strongly disagree to 5 /strongly agree)

The Moral Internalization Scale has been used among adolescents and young adults in the general population (Aquino & Reed, 2002; Hardy et al., 2014; Hardy et al., 2015) and has shown favorable reliability (α = 0.80) and construct validity (e.g., Aquino & Reed, 2002). The scale demonstrated reasonable internal consistency in the current multi-state sample as well (α = 0.79). The mean score of moral internalization among non-referred youth (e.g., Hardy et al., 2015) is higher than that in the current sample (see Table 4 in Appendix B).

I use each youth's average or sum of scores of the Moral Internalization Scale as a marker of their prosocial identity prominence. Depending on the analytical approach, prosocial

[3] In descriptive analyses such as calculating the mean or median, average scores from items were used to align with the original scale of the measure. The exception was general self-efficacy and school connectedness, for which sum scores have always been used in the literature, and thus were used here. For other analyses such as regression or correlation where the variable was treated as continuous, sum scores of items were used to preserve maximal information. This approach was applied to other baseline measures of independent variables.

identity prominence can be a continuous variable based on the original scores, or a dichotomous variable with two levels (low/high) based on 50% percentile.

Prosocial Identity Validation. The PF Survey does not include questions directly measuring prosocial identity validation, which refers to the degree of confidence youths have in their ability to achieve a possible prosocial identity. However, an alternative approach is to use the Generalized Self-Efficacy Scale (Schwarzer & Jerusalem, 1995) in the survey as a proxy of youths' prosocial identity validation status.

Self-efficacy reflects one's confidence in their ability to manage motivation, handle stress, and achieve goals (Bandura, 1977). Studies show that general self-efficacy, which applies broadly, often positively correlates with self-efficacy in specific areas like work, family, or academics (Grether et al., 2017; Schwoerer et al., 2005). Thus, a youth's confidence in developing a prosocial identity, which is a specific kind of self-efficacy, can be approximated by their general self-efficacy when a more precise measure is not available.

The 10-item Generalized Self-Efficacy Scale has been validated across ages 12 to 94 and various cultures, shows good internal consistency (α = 0.75 to 0.94) and construct validity (e.g., Löve et al., 2012; Lönnfjord & Hagquist, 2018; Scholz et al., 2002). The scale demonstrated reasonable internal consistency in the current multi-state sample as well (α = 0.81). Sample items include "*I can always manage to solve difficult problems if I try hard enough*" and "*It is easy for me to stick to my aims and accomplish my goals*" (response scale = 1/not at all true to 4 /exactly true). The international average self-efficacy score based on 19,120 participants across 25 countries was 29.55 (SD=5.32, Scholz et al., 2002). In the current multi-state sample, the mean self-efficacy score was 29.94 (SD=4.67).

Depending on the analytical approach, self-efficacy in the current study can be treated either as a continuous variable based on the original sum of scores, or a dichotomous variable with two levels (low/high) based on 50% percentile.

Social Support

Social support is measured using 9 items from the peer and caregiver clusters of the Child and Youth Resilience Measure (CYRM-28) that assess a youth's perception of the extent of support they receive from peer and caregivers (Ungar & Liebenberg, 2012). Only two items comprise peer support: *"I feel supported by my friends"* and "*My friends stand by me during difficult times*" (response scale = 1/not at all to 5 /a lot). Seven items comprise caregiver support, including, for example, *"My caregivers stand by me during difficult times", and "I talk to my caregivers about how I feel."*

The CYRM-28 has been validated among children and youth with diverse needs and across different cultures (Liebenberg et al., 2012; Ungar & Liebenberg, 2012). In this study, measures of peer and caregiver support showed reasonable reliability (α = 0.85 and 0.86) and factor structure (CFI= 0.97, RMSEA=0.08, applicable only for caregiver support). Given these results, I assessed each youth's perception of peer and/or caregiver support using their average or total scores on relevant measures. These scores can be analyzed as continuous based on the original scores, or dichotomous with two levels (low/high) based on 50% percentile.

School Connectedness

School connectedness is measured by the 5-item School Connectedness Scale that has shown good reliability (α = 0.82 to 0.88) and concurrent validity (*r* = 0.44 to 0.55) across youth groups in different sociocultural contexts (Furlong et al., 2011). The scale demonstrated reasonable internal consistency in the current multi-state sample as well (α = 0.84). Sample items include *"I feel close to people at this school"*, *"I feel like I am part of this school", "The teachers at this school treat students fairly"* (response scale = 1/strongly disagree to 5/strongly agree). I used each youth's average or sum of scores from the scale as an assessment of their degree of school connectedness, which can be a continuous variable based on the original scores, or a dichotomous variable with two levels (low/high) based on 50% percentile.

The above baseline measures exhibit correlations with each other and with relevant constructs in a manner consistent with the literature and their theoretical relations (see Appendix B for details).

Reoffending

Reoffending or recidivism can be measured by two different markers: new court filings or petitions and new convictions. New petitions provide a closer measure of youths' actual delinquent behavior, whereas convictions, though important, may introduce selection bias due to the response of the system. Therefore, I used any new petition as the measure of reoffending, where the presence of any new petition within one-year after the start date of the recidivism follow-up was assigned a value of 1, and the absence of new petition was assigned a value of 0. Here the "start date" is defined as (a) the date of taking the PF survey *or* the date of administering the first risk-need assessment after the baseline referral, whichever comes first; (b) the date of taking PF survey, if no risk-need assessment record. As previously mentioned, recidivism data are currently available only from the PA juvenile justice system, with those from the criminal justice system missing. Therefore, I narrowed the PA subsample to youths under age 17 to ensure that the one-year recidivism records were complete.

Baseline Covariates

Age. Age is an important covariate, as it is related to both psychosocial developments including identity maturity (Hardy et al., 2014; Krettenauer, 2011; Pratt et al., 2003) *and* the ebb and flow of delinquent behavior during adolescence (Piquero et al., 2003; Scott & Steinberg, 2008). In the study, the age of the youth on the start date of the recidivism follow-up is used as a measure of age.

Gender. Previous studies have found that female adolescents exhibit higher levels of moral ideal self than males, but the links between moral ideal self and behavioral outcomes do not differ across genders (Hardy et al., 2013; Hardy et al., 2014). Given the potential gender differences in both prosocial identity and delinquent behavior, gender is considered as a covariate and coded as male or female in the study.

Race. Racial identity is an important social identity. The stigmatized associations between important social identities and negatively stereotyped images can foster a belief that a possible prosocial identity is "not for people like me" (Oyserman et al., 2006). For instance,

African American boys are more likely to be labeled as "troublemakers" (Okonofua & Eberhardt, 2015), and may feel that "people like me" do not belong or cannot succeed in school, especially when aware of being perceived through negative stereotypes (Murphy & Taylor, 2012; Steele, 2010). Additionally, youth from racial minority groups are disproportionately represented in arrest, conviction, and incarceration statistics. Therefore, racial identity is considered as a covariate and coded as four categories, White, Latino, Black, and other/unsure, in the study.

Socioeconomic Status. Socioeconomic status (SES) is associated with both identity development and juvenile delinquency for adolescents. Theory and empirical findings suggest that lower SES correlates with derogatory self-relevant information, poor opportunity structure, and excessive stress (Phillips & Pittman, 2003; Rivnyák et al., 2022), all of which can impact adolescents' identity formation. Lower family SES has been identified as a risk factor for juvenile delinquency as well (Lösel & Farrington, 2012). SES in the study is measured by the Social Deprivation Index (SDI), which is a composite measure of area level deprivation based on demographic characteristics, such as percent living in poverty, percent with less than 12 years of education, and percent single-parent households, that are collected in the American Community Survey by the United States Census Bureau (Butler et al., 2013; Social Deprivation Index (SDI), 2021). Since its development, SDI has been used in multiple published studies (Social Deprivation Index (SDI), 2021). In the current study, I used the SDI by zip code, which were calculated from the ASC (2015-2019) data. Each youth in the sample has an SDI score based on the zip code of their residential address. The SDI scores take values from 1 to 100, with a higher value indicating a higher level of social deprivation.

Aggregate Risk of Reoffending. In this study, youths typically undergo a risk-need assessment at the intake of community supervision. As such, scores from this assessment—often coinciding with when the youths took the PF survey—can be used to measure their baseline risk (of recidivism). The PA juvenile justice system employs the Youth Level of Service/Case Management Inventory Version 2.0 (YLS/CMI) (Hoge & Andrews, 2011) as its assessment tool, where youth's levels of risk are represented by scores ranging from 0 to 32.

Incremental Approach to Adjust for Covariates. There are both pros and cons, for including covariates like socioeconomic status and baseline risk in analyses. On the one hand, risk scores or levels generated from comprehensive risk-need assessment tools are strong predictors of reoffending. In the current study, the point-biserial correlation between baseline risk score and one-year recidivism is 0.22. On the other hand, risk scores are aggregated measures of recidivism risk that span multiple domains such as family, peer, education, attitude, aggression tendency, personality, and substance use (Hoge & Andrews, 2011). These domains are likely to overlap conceptually and empirically with the prosocial identity and allied measures of interest (e.g., caregiver support). In this study, among the PA subsample, prosocial identity measures and the proposed moderators are inversely correlated with baseline risk (Kendall's τ ranges from -0.10 to -0.20). Adjusting for baseline risk may help clarify the unique association between prosocial identity and delinquent outcomes in a stringent manner, but it may also mask real effects by statistically suppressing the potential shared variance among the predictors (i.e., prosocial identity measures and moderators), the covariate (i.e., baseline risk), and the outcome (i.e., recidivism). If the covariate is a potential mediator, which remains an

open empirical question in this case, then adjusting for it could block the real effect between the predictor and the outcome (Meehl, 1971; Pearl, 2010). Given these concerns, covariates such as age, gender, race, SDI, and baseline risk are added stepwise in regression models to observe their influence on the relations among variables (see more details in the next section, "Analytical Approach").

Summary of Measure Design. Figure 2 shows an overview of the study design, data source, measures, and variables. Details of measures at the item level for baseline variables and other constructs (that will be involved in psychometric analyses) are provided in Appendix A.

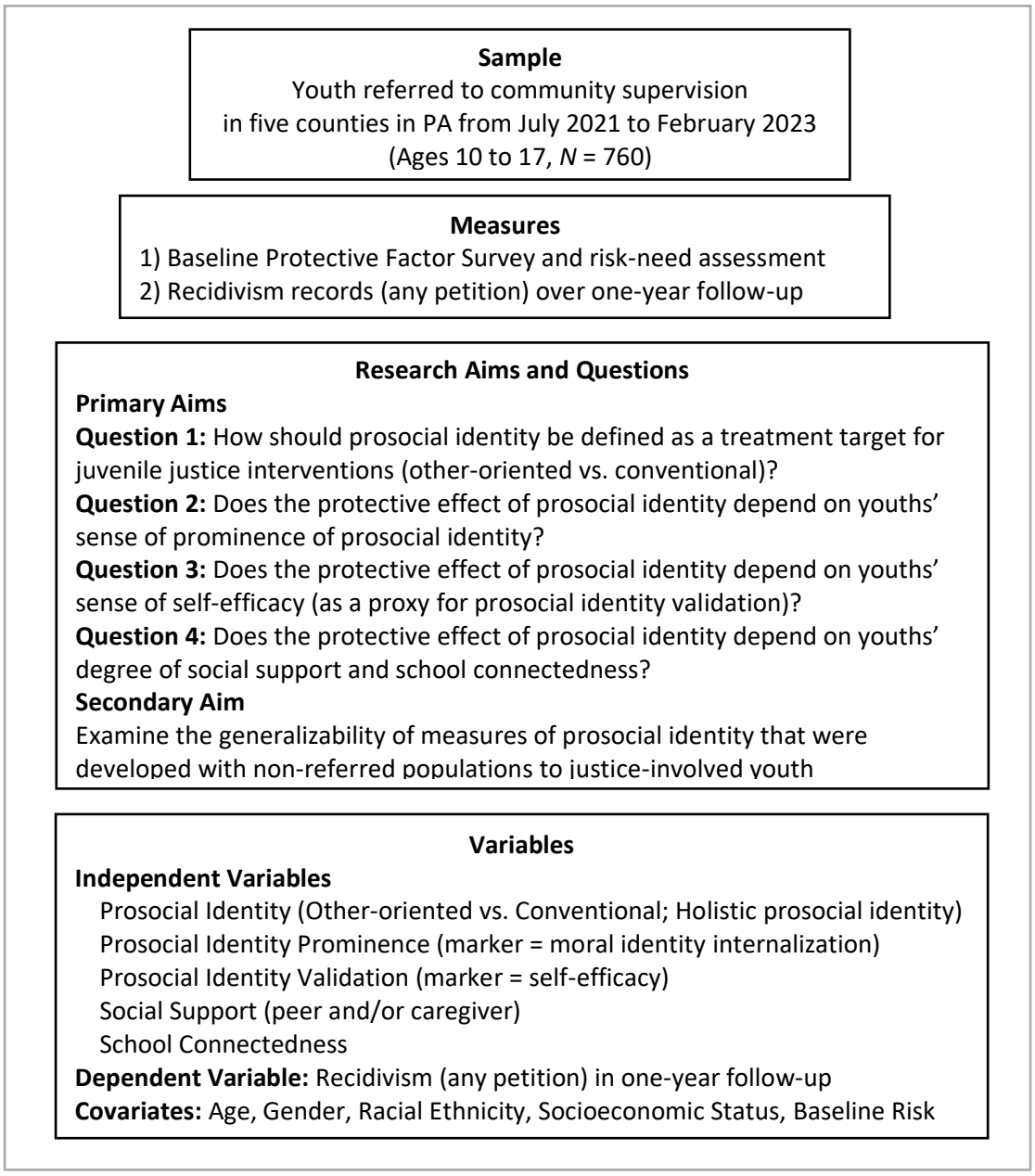

Figure 2. Study Design

Analytic Approach

In the results section below, I first provide descriptive statistics and preliminary analyses of basic relations among variables. I then proceed to the primary analyses to address the four research questions. The primary analyses adopt an incremental approach—starting with no covariates and then gradually layering in covariate sets and using increasingly stringent analytic approaches (from logistic regressions to causal modeling). As noted earlier (see "incremental approach" above), there are both pros and cons to statistically adjusting for complex variables that overlap with the protective factors of interest. After explaining how missing data were handled, I provide detail on descriptive and primary analyses.

Preliminary analyses. Preliminary analyses focus on describing the study sample; the relations among measures of prosocial identity, proposed moderators, and covariates; and relations between measures of protective factors and the recidivism outcome. The methods used include basic descriptive statistics, bivariate correlations, univariate logistic regressions, and multiple logistic regressions with covariates being added stepwise. These preliminary analyses serve to build a foundation for the next steps by understanding the data structure, identifying significant relations, exploring the influence of covariates, and establishing a baseline for comparison for more sophisticated analyses.

Primary analyses. To address question one (i.e., how prosocial identity should be defined as a treatment target for juvenile justice interventions), logistic regressions with other-oriented and conventional identity being added stepwise are used to examine the incremental utility of each of the two variables in predicting recidivism.

To address questions two to four (i.e., whether the protective effect of prosocial identity depend on youths' sense of prominence of prosocial identity, self-efficacy, social support, and school connectedness), three analytical approaches were used:

(a) Multiple logistic regressions with interaction terms, initially without adjusting for any covariates, and then adding covariates stepwise if interaction effects are found. Incorporating interaction terms in regression models is a conventional approach to detect moderation effects between variables. As mentioned earlier, adjusting for covariates can reveal the unique protective effects of prosocial identity on recidivism—but can also obscure the real effects when "over-adjusted" by including nuanced variables such as baseline risk. Therefore, I use an incremental approach to add covariates to the regression models.

(b) Targeted Maximum Likelihood Estimation (TMLE) for the main effects of holistic prosocial identity on recidivism. As a relatively new approach that merges the strengths of the Inverse Probability of Treatment Weighted (IPTW) method (Robins et al., 2000) with machine learning techniques, TMLE is a double-robust, semi-parametric estimation procedure that approximates experimental conditions by creating comparison groups that are similar in their observed characteristics. This approach takes account for the potential confounding effects from observed covariates (i.e., age, gender, race, SDI, and baseline risk) to estimate the effect of the variable of interest (i.e., prosocial identity measured by moral ideal self) on the outcome (i.e., recidivism). Some advantages of TMLE over traditional regression methods include that TMLE is less dependent on the specification of the outcome model; it can achieve a better bias-

variance trade-off in estimates; it handles confounding more effectively by incorporating the IPTW method; and it offers flexibility and potentially improved performance by incorporating machine learning algorithms to model complex relations between variables. For more details on TMLE procedures, see van der Laan & Rose (2011). Applying TMLE to estimate the main effects of prosocial identity on recidivism can further verify findings from the preliminary analyses if significant direct relations (as opposed to moderating effects) are identified. However, because aggregated measures such as SDI and baseline risk are included as covariates, the issue of "over-adjusting" for confounders applies with equal or more force to TMLE. Methodological limits (i.e., not manipulating prosocial identity) cannot be entirely solved by statistical approaches.

(c) Marginal Structural Modeling (MSM) with an effect modifier, using the TMLE estimator (Robins et al., 2000; van der Laan & Rose 2011). This approach serves the same purpose as logistic regression with interaction terms (a, above) in detecting the proposed moderation effects. The differences are that (c) represents a causal analytical approach and incorporates the strengths of TMLE to estimate moderation effects in a more rigorous manner. If moderation effects are detected by conventional regression models, then MSM with the TMLE estimator can further verify if these effects hold under a more stringent analytical approach. Meanwhile, adopting a causal framework can estimate how the expected probability of reoffending vary as a function of youth's level of prosocial identity and level of the proposed moderators. Interpretations of these estimates may be associational rather than causal, given that the assumptions for casual inference are rarely met in observational data. However, the advantage of employing a formal causal framework lies in its ability to "help in designing a statistical analysis that comes as close as possible to answering the motivating causal question, while making clear what assumptions are required to endow the resulting estimates" (Petersen & van der Laan, 2014, p. 1). Therefore, even if the results from the analyses cannot be interpreted as causal, they at least offer insights into how different conditions of interest are associated with varying probabilities of the outcome (i.e., reoffending) occurring.

Please see Appendix C for further details, including conceptualizations of background knowledge on variables and model specifications for both the regression models and MSMs with the TMLE estimator.

Missing data. There were very few missing data for independent variables, covariates, and the outcome. In fact, only the baseline risk score, SDI score, and the Moral Internalization Scale had missing rates of 5.5%, 0.1%, and 0.4%, respectively. For analyses involving TMLE, I imputed the sample's median for continuous variables (e.g., risk score) and used resampling with replacement of the non-missing values for categorical variables (e.g., when moral internalization was coded as a binary variable with low or high level). I also included a missingness indicator for each of the covariates (i.e., risk score and SDI score) in analyses involving TMLE. For other analyses, such as regression and correlation, a listwise deletion approach was used, where rows with missing data were dropped.

Data preparation and computational work involved in the analyses were completed in R (version 4.0.3) and performed in a secure remote workspace hosted by the Secure Research, Data, and Compute service team at UC Berkeley. Specifically, for the TMLE and MSM with TMLE

analyses, I used the “ltmle” package (Schwab et al., 2020), and the “SuperLearner” package (Polley et al., 2021). SuperLearner candidates are the default algorithms (i.e., SL.glm, SL.step, and SL.glm.interaction) in the library. Both preliminary and the main analyses are based on data from the PA subsample of 760 youths, as defined earlier.

Results

Preliminary Analyses

Descriptive Statistics and Associations Among Baseline Characteristics

Table 2 provides descriptive statistics for measures of prosocial identity and proposed moderators. The Pennsylvania sample's descriptive statistics closely align with those of the larger multi-state sample. Each scale used in primary analyses also exhibits reasonable internal consistency.

Table 2. Descriptive statistics for measures of independent variables (*N*=760)

Variable	Measure	Min.	Median	Mean	Max.	SD	Cronbach's α
Prosocial Identity Content (holistic)	Moral Ideal Self Scale (MISS)	1	6	5.88	7	0.85	0.93
Other-oriented Identity	Loving/Caring cluster of MISS	1	6.11	5.96	7	0.88	0.87
Conventional Identity	Other four clusters of MISS	1	5.91	5.82	7	0.92	0.91
Prosocial Identity Prominence	Moral Internalization Scale	1.6	4	3.98	5	0.7	0.81
Prosocial Identity Validation	Generalized Self-Efficacy Scale	14	30	29.87	40	4.5	0.81
Social Support	CYRM-28 (Peer & Caregiver)	1.13	4.22	4.09	5	0.7	0.83
	CYRM-28 (Peer)	1	4	3.93	5	1.08	0.86
	CYRM-28 (Caregiver)	1.14	4.29	4.14	5	0.76	0.85
School Connectedness	School Connectedness Scale	5	17	16.87	25	4.47	0.84

As shown in Table 3, measures of other-oriented and conventional identity are strongly correlated (τ= 0.58, $p < 0.001$). Consistent with the literature, the Moral Ideal Self Scale moderately correlates with moral internalization (τ= 0.43 or 0.36, $p < 0.001$). Other pairwise correlations are relatively weak, indicating good divergence among constructs. However, social support—particularly caregiver support—is moderately associated with the Moral Ideal Self Scale (τ= 0.31-0.34), and moral internalization (τ= 0.28).

Table 3. Bivariate correlations among independent variables (*N*=760)

	Moral Ideal Self	Other-oriented Identity	Convent-ional Identity	Moral Internali-zation	Self-efficacy	Social Support	Peer Support	Caregiver Support	School Connected-ness
Moral Ideal Self	1								
Other-oriented Identity	0.77***	1							
Conventional Identity	0.83***	0.58***	1						
Moral Internalization	0.43***	0.43***	0.38***	1					
Self-efficacy	0.24***	0.20***	0.25***	0.13***	1				
Social support - peer & caregiver	0.34***	0.32***	0.31***	0.26***	0.28***	1			
Peer Support	0.18***	0.17***	0.16***	0.13***	0.23***	0.48***	1		
Caregiver Support	0.34***	0.33***	0.31***	0.28***	0.25***	0.81***	0.23***	1	
School Connectedness	0.21***	0.20***	0.19***	0.19***	0.19***	0.31***	0.22***	0.28***	1

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

Correlations were also computed for the independent variables and covariates (see Appendix D for details). Results indicate that some of the independent variables are weakly associated with age and socioeconomic status (Kendall's τ ranges from -0.10 to 0.06). More importantly, many of the independent variables were weakly to moderately associated with aggregate recidivism risk scores (Kendall's τ ranges from -0.10 to -0.20). This is expected, as all the independent variables are theoretically protective factors against reoffending.

Bivariate Associations Between Baseline Characteristics and Recidivism

Table 4 shows bivariate associations between baseline characteristics and recidivism. For correlations between continuous variables and the binary recidivism outcome, point-biserial correlation was applied, with Pearson's product-moment correlation as the test statistic; for correlations between categorical or dichotomous variables and the binary recidivism outcome, the Chi-square test was used to examine if levels in the categorical variable differed between those who reoffended and those who did not.

Results indicate that most baseline covariates significantly predict reoffending. Among the prosocial identity measures and proposed moderators, **other-oriented identity**, **social support**, and **school connectedness** weakly protect against recidivism (r = -0.07 to -0.11).

Table 4. Association between baseline characteristics and recidivism outcome (*N*=760)

		Any Petition within One Year	
		r_{pb}	χ^2
Baseline Covariates and Characteristics	Age at intake	0.07*	
	Gender (male/female)		4.14*
	Race (White, Latinx, Black, other/unsure)		27.50***
	SES (Social Deprivation Index score)	0.14**	
	YLS total risk scores	0.22***	
	Type of current referral (4 categories)		1.09
	Current referral being violent		0.01
	Number of prior offenses	0.07	
Main Predictors	Moral Ideal Self	-0.05	
	Other-oriented Identity	-0.08*	
	Conventional Identity	-0.02	
Potential Moderators	Moral Internalization	-0.07	
	Self-efficacy	-0.02	
	Social Support (Peer & Caregiver)	-0.11**	
	Peer Support	-0.05	
	Caregiver Support	-0.11**	
	School Connectedness	-0.07*	

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Adjusted Associations Between Prosocial Identity, Proposed Moderators—and Recidivism

To explore how the predictive power of independent variables on recidivism may change in the presence of different covariates, a series of multiple logistic regressions were conducted. Each regression model used prosocial identity measure or a proposed moderator as the main predictor. For each main predictor:

1) Initially, basic demographics including age at intake, gender, and race were added as covariates in the models.
2) Then, the socioeconomic measure (SDI score) was included as an additional covariate.
3) Finally, the baseline risk score was added as another covariate.

Results indicate that the predictors of interest, including other-oriented identity, do not significantly predict recidivism after adjusting for even basic demographic covariates (#1 above). The only exception is for one proposed moderator. Social support—including caregiver support (OR=0.95, $p < 0.01$) and the combination of peer and caregiver support (OR = 0.96, $p < 0.05$)—significantly predicts recidivism after adjusting for demographics and SDI scores (#1 & 2 above), but not aggregate risk scores (#3 above). For details, see Table 4 in Appendix D.

Summary of Relevant Results from Preliminary Analyses

In summary, two relevant messages emerge from the preliminary analyses. First, other-oriented identity is the only prosocial identity measure that significantly predicts recidivism, although some potential moderator variables are also predictive (i.e., social support and school connectedness). Second, after adjusting for a range of covariates, only *social support* remains a weak but significant protective factor against recidivism (OR = 0.95 to 0.96)—which drops to insignificance when the covariates include aggregate risk scores.

Primary Analyses

Finding 1: Unlike Conventional Identity, Other-Oriented Identity Protects Against Reoffending

My first hypothesis was that conventional identity would predict offending more strongly than other-oriented identity—and the latter variable would add no value to the former variable in predicting recidivism. This hypothesis was not supported. Instead, I found that only other-oriented identity predicted recidivism in these initial analyses.

Table 5 presents results from comparing three regression models that use 1) other-oriented identity as the single predictor, 2) conventional identity as the single predictor, and 3) both other-oriented and conventional identity as predictors of the recidivism outcome, respectively. The results indicate that other-oriented identity significantly predicts recidivism, both as a single predictor and when added to the model alongside conventional identity. Conversely, conventional identity alone, or in combination with other-oriented identity, does not significantly predict recidivism. Both the Akaike Information Criterion (AIC) and Likelihood Ratio Tests suggest that adding conventional identity does not improve model fit. In other words, adding other-oriented identity alongside conventional identity improves the prediction of youth recidivism within one year, but not vice versa. Specifically, a one-unit increase in the

other-oriented identity measure is associated with a 5% decrease in the odds of recidivism within one year, holding conventional identity constant.

Table 5. Other-oriented identity, but not conventional identity, predicts recidivism (*N*=760)

	Predictor	Any Petition within One Year			
		OR	95% CI	AIC	Likelihood Ratio Test
Model 1	Other-oriented Identity	0.97*	0.95, 1.00	610	
	Intercept	0.72	0.19, 2.56		
Model 2	Conventional Identity	0.99	0.98, 1.02	614	
	Intercept	0.22*	0.06, 0.79		
Model 3	Other-oriented Identity	0.95**	0.91, 0.98	608	Model 3 vs. Model 1 Model 3 vs. Model 2**
	Conventional Identity	1.03	1.00, 1.06		
	Intercept	0.49	0.12, 1.87		

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Finding 2: No Evidence that Prosocial Identity Protects Against Recidivism after Rigorously Adjusting for Covariates or Using Causal Modeling

As indicated earlier (see preliminary findings), other-oriented identity does not significantly predict recidivism after adjusting for even basic covariates. Moreover, prosocial identity measured holistically (as total MISS scores) did not significantly predict recidivism (see Table 4). However, because total MISS scores were used to test moderation questions, I used TMLE to estimate its causal association with recidivism.

Not surprisingly (given null bivariate effects), Targeted Maximum Likelihood Estimation (TMLE) provided no evidence that prosocial identity (total MISS scores, point estimate = -0.03, 95% CI: -0.07 to 0.02) significantly protected against recidivism. Notably, to conduct TMLE analyses, total MIS scores were coded as binary (low/high) based on the 50th percentile score. Covariates include age at intake, gender, race, SDI score, and baseline risk. The assessment of the positivity assumption—a key assumption of the TMLE estimator—was done by evaluating the distribution of the estimated propensity scores for participants to be in each level of the holistic prosocial identity, conditioned on the covariates. Propensity scores were estimated using logistic regression models. The results indicate that the propensity scores for the dichotomous holistic prosocial identity ranged from 0.13 to 0.78, and the density plot of the propensity scores had a reasonable inverted U-shape without extreme values (e.g., less than 0.05 or greater than 0.95). This suggests that there was no violation of the positivity assumption for the main predictor (dichotomous holistic prosocial identity).

Finding 3: Peer Support Significantly Moderates the Relation Between Prosocial Identity and Recidivism

My hypothesis for questions two to four was that the protective effects of prosocial identity would depend on youth's sense of prominence of prosocial identity, self-efficacy, social support, and school connectedness. Although the main effect of prosocial identity on recidivism reduction is not significant, testing for interaction effects is worthwhile because an interaction effect may still exist and reveal important conditional relations that are not evident when considering the main predictor alone (Lorah, 2020; Zhang, 2020).

This hypothesis for questions two to four was partially and weakly supported. Among the proposed moderators, peer support significantly moderated the relation between prosocial identity and recidivism in logistic regressions even when adjusting for all the covariates. This moderation effect was not detected after adopting a causal analytic approach. Details are presented below.

Regression Analysis to Test Moderation Effects: Peer Support Moderates the Effect of Prosocial Identity on Recidivism, Using Rigorously Adjusted Variables. As proposed, multiple logistic regressions with interaction terms were conducted to examine the extent to which the sense of prosocial identity prominence (i.e., moral internalization), general self-efficacy, social support (from peers and/or caregivers), and school connectedness moderate the relation between prosocial identity (i.e., moral ideal self) and recidivism. These regressions began without any covariates. If statistically significant moderation effects were found, covariates including age, gender, race, SDI score, and baseline risk score were then added to the model stepwise to observe whether the moderation effects would persist.

The results indicate that there was only one interaction effect: **moral ideal self** interacted with **peer support** (OR = 0.99, 95% CI: 0.99 to 1.00, $p < 0.05$) to predict recidivism. **This interaction effect holds even after adding <u>all</u> covariates including risk estimates into the model** (OR = 0.99, 95% CI: 0.99 to 1.00, $p < 0.05$). Results are shown in Table 6—both when no covariate was adjusted and when all covariates were added to the model, respectively.

Table 6. Moral ideal self interacts with peer support to predict recidivism (*N* = 760)

Variable	Any Petition within One Year			
	OR	95% CI	OR	95% CI
PI Content (Moral Ideal Self)	1.04*	1.00, 1.08	1.04*	1.00, 1.08
Peer Support	2.06*	1.15, 3.86	1.85	1.02, 3.53
PI Content x Peer Support	0.99*	0.99, 1.00	0.99*	0.99, 1.00
Age at Intake			1.15	0.98, 1.37
Female			0.62	0.36, 1.03
Race-Latinx			1.01	0.41, 2.36
Race-Black			2.15*	1.20, 3.96
Race-Other/unsure			0.58	0.09, 2.25
SDI Score			1.01	1.00, 1.01
Baseline Risk Score			1.08***	1.04, 1.11
Intercept	0.00**	0.00, 0.17	0.00***	0.00, 0.01

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Figure 3 illustrates how peer support moderates the effect of moral ideal self on recidivism, using the model without covariates in Table 7. The "Low," "Moderate," and "High" levels of moral ideal self and peer support were determined by dividing the sum of scores for the measures of these two variables into tertiles. The graph shows that the protective effects of moral ideal self are observed only for those with moderate to high levels of peer support. If one does not feel supported by peers (i.e., perceived peer support is low), moral ideal self does not predict a decrease in recidivism. In fact, when moral ideal self is low, higher peer support is associated with a slightly higher predicted probability of reoffending.

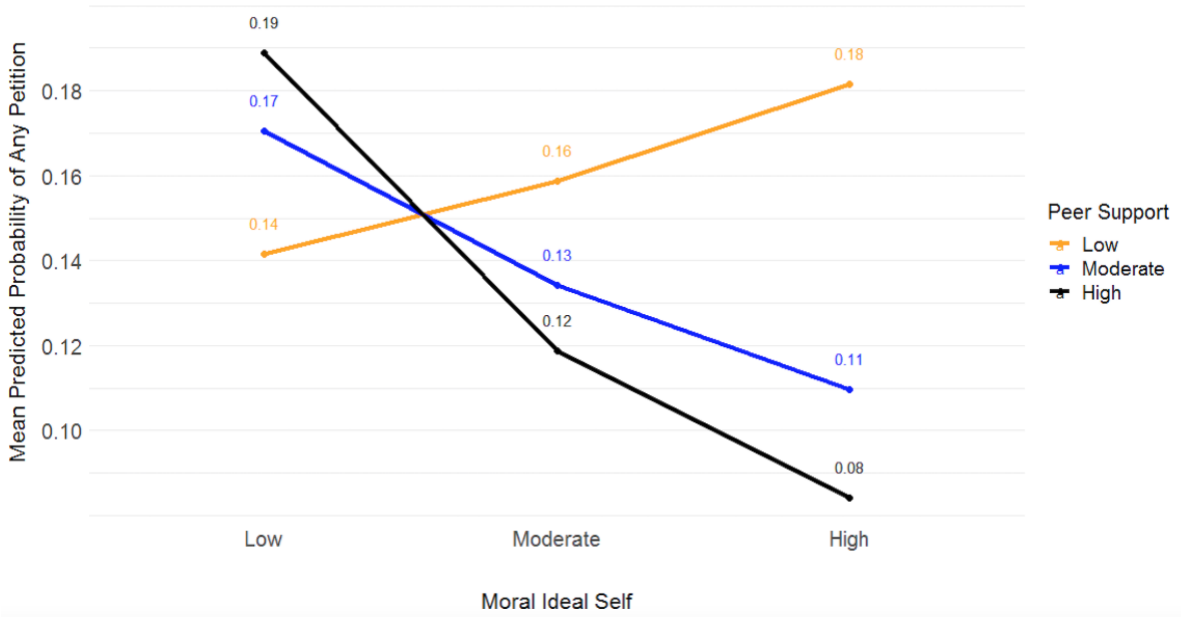

Figure 3. Peer support Moderates the Effects of Prosocial Identity (Moral Ideal Self) on Recidivism

Marginal Structural Models (MSM) with Effect Modifier using TMLE Estimator: No Moderation Effect was Detected after Adopting a Causal Analytic Approach. As an alternative way to test the hypothesized moderation effects, MSM with effect modifiers and TMLE estimator were conducted to examine whether the expected probability of reoffending varies as a function of youths' level of prosocial identity and levels of the proposed moderators, taking the potential confounding effects from covariates into account. Results indicate that no significant interaction effects.

Summary of Results from Primary Analyses

The primary analyses for addressing the four research questions of the study yield several findings. *First,* **other-oriented prosocial identity, but not conventional identity, predicts recidivism** independently. These results contradict my hypothesis that conventional (but not other-oriented) identity would uniquely predict recidivism.

Second, **after rigorously adjusting for covariates and/or adopting a causal analytic approach, prosocial identity (as a total score, combining other-oriented and conventional identity) does not protect against recidivism**.

Third, **peer support significantly moderates the relation between moral ideal self and recidivism (OR=0.99)** in logistic regression analyses, even when adjusting for *all* baseline covariates. Counter to my hypotheses, there were null moderation effects for the remaining moderators of interest (e.g., moral internalization, self-efficacy, school connectedness).

Fourth, **after adopting a causal analytic approach, peer support no longer significantly moderates the relation between moral ideal self and recidivism.** Counter to my hypotheses, MSMs with TMLE analyses indicate no significant moderation of the effect of moral identity on recidivism for any target variables.

These findings collectively suggest that other-oriented identity plays a more critical role than conventional identity in predicting reoffending when not adjusting for other covariates. Although interaction effects among the protective factors of interest in this study are minimal, peer support moderates the effects of prosocial identity (i.e., moral ideal self) on recidivism, even after adjusting for covariates including race and baseline risk.

Discussion

In this study I used measures of prosocial identity validated in the general youth population, along with other individual and environmental protective factors, to explore their impact on recidivism outcomes among a sample of justice-involved youth. The Positive Youth Development (PYD) philosophy and its extension to the juvenile justice context—Positive Youth Justice (PYJ)—have been advocated by researchers and practitioners. However, there is a limited understanding of how the positive constructs within the PYD framework protect youth against reoffending, not to mention empirical guidance on translating the PYJ approach into real-world interventions. In this study, I focus on prosocial identity, one promising construct within the PYD framework, to empirically examine theoretical propositions from a conceptual model (Jian & Skeem, 2023) that specifies one possible mechanisms of behavioral change among justice-involved youth. The central inquiries include *how* prosocial identity should be defined as a treatment target in the juvenile justice context and under *what conditions* prosocial identity is protective against reoffending.

To address these questions, I first examined the psychometric properties of prosocial identity measures in a sample of 1,913 court-referred youths from eleven counties across three states (Pennsylvania, Virginia, and Wisconsin). Then, I conducted a series of progressively stringent analyses, including traditional regression and causal modeling approaches that incorporate new developments in computational social science, using data from a subsample of 760 youths (ages 10 to 17) who were referred to community supervision in five counties in Pennsylvania (PA) between July 2021 and February 2023.

The principal results suggest that 1) other-oriented identity appears more important than conventional identity in predicting recidivism; 2) prosocial identity (as a combination of other-oriented and conventional identities) does not significantly protect against recidivism after rigorously adjusting for covariates and/or adopting a causal analytic approach; and 3) only peer support significantly moderates the relation between prosocial identity and recidivism (and does so even when adjusting for all covariates; though not when adopting causal analyses). Additionally, psychometric analyses indicate that prosocial identity measures, such as the Moral Ideal Self Scale and Moral Internalization Scale developed with non-referred youth, can be generalized to justice-involved youth. In this section, I unpack each finding below and reflect on the study's strengths and limitations.

The Main Predictive Value of Other-Oriented vs. Conventional Identities, and Holistic Prosocial Identity, on Recidivism

The first aim of the study is to examine the basic protective effect of other-oriented vs. conventional identities, as well as holistic prosocial identity as a combination of the two, in a justice-involved sample. Contrary to my hypotheses, I found that other-oriented identity more important than conventional identity in predicting recidivism, and that no empirical support that the holistic prosocial identity measured by Hardy's Moral Ideal Self scale (MISS) protects against recidivism.

Other-Oriented Identity, but not Conventional Identity, Predicts Recidivism

This finding contradicts my initial hypothesis that conventional identity, as often emphasized in criminological literature, would be a more important predictor of recidivism among justice-involved youth. Nonetheless, this result seems to align with findings from some adjacent studies that explore the relation between moral development and recidivism. Moral development is a broad concept encompassing moral judgment (i.e., justifying decisions or values based on considerations of fairness, right, and wrong), empathy (e.g., understanding or sharing another's emotional state), and moral emotions (e.g., guilt or shame) (see the review by Van Vugt et al., 2011). While these internal processes are not explicitly identified as moral or other-oriented identity, they are likely relevant to one's moral self-images featured by an other-orientation. In fact, "most developmental scientists agree that morality refers to individuals' treatment of others" and provides motivation for prosocial behavior (Killen & Smetana, 2015, p.702-703). A meta-analysis of 19 studies (N = 15,992) found a small but significant inverse relation (r = 0.10) between moral development and recidivism (including both self-reported and officially recorded) among justice-involved youth (Van Vugt et al., 2011). Together with the finding from the current study, it appears that moral functioning—including an other-oriented identity—can be a protective factor against recidivism in the juvenile justice context.

This finding implies that juvenile justice interventions aimed at fostering prosocial identity should prioritize cultivating motivations and values that encourage a desire to benefit others, alongside promoting prosocial behaviors like caring, helping, sharing, and contributing. This approach sets a higher standard than merely pursuing educational or vocational goals or adhering to laws and social norms.

However, the correlation between other-oriented identity and recidivism, while significant, is not strong. Thus, this finding should be interpreted with caution. Although small effect sizes are common when predicting complex behaviors such as delinquency with a single predictor (Ahadi & Diener, 1989), given that this study represents an initial exploration into the comparative predictive value of other-oriented versus conventional identities on recidivism, further research is needed.

To validate this finding, future research can improve in two ways: 1) by using diverse samples of justice-involved youth that encompass various characteristics, especially risk levels, and 2) by using more precise instruments capable of distinctly differentiating between other-oriented and conventional identities as described in the conceptual model. Although psychometric analyses indicate that the two concepts can be differentiated using the MISS, the scale was primarily developed to measure moral identity content—which is at the higher end of the prosocial identity spectrum (i.e., leaning towards an other-orientation). There is room to refine the conventional identity measure, for example, by reflecting specific hoped-for selves who achieve educational and vocational goals (rather than abstract conventional traits one desires to have), to align more closely with the conceptualization of conventionally featured prosocial identity in the criminological literature.

No Empirical Support that Prosocial Identity Protects Against Recidivism

The study employed multiple analytical approaches, including logistic regressions and Targeted Maximum Likelihood Estimation (TMLE), to examine the basic effects of holistic prosocial identity measured by moral ideal self, and prosocial identity prominence measured by moral internalization, on recidivism among justice-involved youth. Contrary to my hypothesis, I found no empirical support that these measures protect against recidivism. This is surprising, given that previous research, though limited, suggests that the predictive utility of prosocial identity on antisocial behavior seems to generalize across measurements and youth populations. Below, I compare the current study with the studies reviewed in the "Introduction" section (see "Evidence that Prosocial Identity Protects Youth Against Antisocial Behavior") to interpret the null effect.

Comparing to Studies Among Non-Referred Youth Using Validated Measures. This group of studies comes from the developmental literature, employing validated measures of moral identity similar to those used in the current study (i.e., the MISS and the Moral Internalization Scale). However, the outcome measures in these studies often focus on externalizing behaviors such as aggression, rule-breaking, and cheating (e.g., Hardy et al., 2014; Hardy et al., 2015), which are less severe than behaviors typically classified as delinquent or criminal. The samples in these studies are predominantly White youth with an even gender distribution, differing significantly from those in the current study, which primarily involves racially minority, male, and economically disadvantaged youth. Therefore, although the bivariate correlations between moral identity measures and externalizing behaviors in those studies are moderate (ranging from -0.14 to -0.33, Hardy et al., 2014), it is expected that this relation would be weakened when using juvenile justice outcomes—like those in the current study—that indicate much more serious delinquent behavior that reaches the threshold of system processing.

Comparing to Studies Among Non-Referred Youth Using Constructed Measures. Rather than using validated measures of moral identity, this group of studies, represented by Na & Paternoster (2019) and Rocque et al. (2016), constructs prosocial identity measures by using items that capture an identity theoretically relevant to crime. These items include questions asking youth the extent to which they viewed themselves as a "troublemaker" and "delinquent," or as a conventionally "good person" (e.g., being law-abiding, trustworthy, honest, and not mean). Outcome measures are self-reported violent or delinquent behaviors (e.g., bullying and threatening, physical assault, gang fights, armed robbery, vandalism, breaking and entering, major theft) that are closer to those in the current study.

Notably, these studies employed a longitudinal design that followed participants for five or more years and found significant inverse relations between prosocial identity and delinquent outcomes—even after adjusting for demographic and theoretically relevant factors (e.g., SES, social bonds). These findings are the reason I initially hypothesized that conventional identity would be a more important predictor than other-oriented identity for recidivism in the current study. Again, the samples in these studies were drawn from the general population, which differ from justice-involved youth.

Without more research, it is difficult to infer potential reasons that may explain the differences between findings from these studies and the current one. Future research can benefit from employing both validated prosocial identity measures (i.e., for moral identity, which leans toward the other-oriented side) and constructed prosocial identity measures (i.e., for conventional identity, like those in the Rocque et al. (2016) study), to compare their effects on recidivism reduction among justice-involved youth.

Comparing to Studies Among Justice-Involved Youth Using Proxy Measures from Risk Assessment. This group of studies, represented by Skeem & Kennealy (2015) and Baglivio et al. (2014), has outcome measures—for example, serious or violent institutional infractions within one year, or petition or conviction within one year after completion of probation—that are close to those in the current study. However, the two studies differ from the current one in two main ways. First, the samples in the two studies, consisting of incarcerated youth and SVC (serious, violent, and chronic) juvenile offenders, are generally at a much higher risk of reoffending than those in the current study. Second, the prosocial identity measures were constructed from the attitude domain of risk assessment tools that more closely align with the "prosocial attitudes" concept. In fact, the Baglivio et al. (2014) study does not have an identity measure—they used the prosocial attitudes domain directly as the predictor. These measures overlap with the conceptualization of conventional identity. For example, the measure of prosocial identity in the Skeem & Kennealy (2015) study attempts to capture participants' intentional self-change toward a prosocial direction by assessing youth's degree of commitment to law-abiding behavior, eagerness to disassociate from a criminal lifestyle, interest in affiliation with prosocial peers, and signs of setting out realistic goals and taking steps to achieve them. Unlike the validated prosocial/moral identity measures that capture youth's sense of a prosocial self at a general level, these measures are more specific to crime (and crime desistance) and have shown predictive utility in recidivism, even for high-risk youth. This might suggest that the conceptualization of conventional identity in future research should focus more on activities directly relevant to crime desistance. In other words, being specific in capturing the hoped-for selves contrary to a (feared) criminal self. This also aligns with research in the risk assessment and psychology literature where more specifically measured or narrowly defined constructs, when closely match theoretical constructs that correlate with outcomes, can improve predictive validity (Hendry et al., 2013; Jenkins & Griffith, 2004).

Overall, the null effect of prosocial identity on recidivism in the current study has three implications. *First,* a more accurate measure of conventional identity—as conceptualized in the criminological literature—among justice-involved youth is needed to clarify its predictive value on recidivism outcomes. As mentioned above, instead of only using general conventional traits (e.g., truthful, respectful, makes good choices) to describe the type of person youth hope to become, future measures should include items that specifically assess hoped-for self-images relevant to crime desistance, such as being law-abiding, disassociating from criminal lifestyles and networks, achieving educational and vocational goals, and fulfilling family and adult responsibilities. This approach also facilitates comparisons with findings from similar studies like Skeem & Kennealy (2015), Na & Paternoster (2019), and Rocque et al. (2016).

Second, the current null effect of prosocial identity does not falsify the propositions of the conceptual model (Jian & Skeem, 2023) that the study aims to examine. This model proposes that prosocial identity, in terms of a hoped-for possible self, would work best to prevent delinquency when balanced by a feared self in the same domain. However, the current study lacks measures on the feared self or the balance with the hoped-for self, thus is not able to fully examine the proposition. Future research can add a supplemental measure of feared antisocial self and assess its balance with the hoped-for prosocial self by asking youth to describe a list of (e.g., three or more) self-images they want to avoid and examine whether these self-images correspond to those they want to realize (i.e., the hoped-for self). The design of such measures can be informed by research on academic possible selves (e.g., Oyserman et al., 2006, Horowiz et al., 2020).

Third, thus far, all the studies in this field relating to prosocial identity and justice-involved youth have been observational. As Kurt Lewin (1943) said, "If you want truly to understand something, try to change it." An experimental design to manipulate or foster aspects of prosocial identity, as suggested by the Jian & Skeem (2023) study, will be an optimal way to test the effect of prosocial identity and the utility of the conceptual model.

Influence of Prosocial Identity Internalization, Self-Efficacy, Social Support, and School Connectedness on the Protective Effect of Prosocial Identity

The second primary aim of the study is to examine whether the proposed moderators—prosocial identity internalization, self-efficacy, peer and caregiver support, and school connectedness—influence the relation between prosocial identity and youth recidivism. Partially consistent with my hypotheses, I found that peer support significantly moderates the impact of prosocial identity on recidivism in the expected direction, even when adjusting for all covariates. This moderation effect was not observed in the causal analytical approach (i.e., MSM with an effect modifier and TMLE estimator). Other than that, neither regression nor MSM analyses detected significant interaction effects between prosocial identity and other proposed moderators.

Peer Support as a Moderator in the Protective Role of Prosocial Identity Against Reoffending

The study found that peer support moderates the relation between prosocial identity, as measured by the moral ideal self, and recidivism. Specifically, the mean predicted probability of any petition within one-year follow-up decreases as the level of moral ideal self increases—but only when perceived peer support is moderate or high. In other words, the protective effects of moral ideal self on preventing reoffending are only observed in those who perceive themselves to have moderate or high levels of peer support. If one does not feel supported by peers (i.e., perceived peer support is low), even a moderate or high level of moral ideal self does not lead to a predicted decrease in recidivism. Additionally, when the level of moral ideal self is low (rather than moderate or high), higher peer support is associated with a slightly higher predicted probability of reoffending.

The finding that prosocial identity has stronger protective effects against reoffending among those with higher levels of peer support is consistent with my hypothesis (although my hypothesis was about general social support, not peer support specifically). The dynamics

among prosocial identity, peer influence, and delinquent behavior have been sparsely explored in previous research. However, indirect evidence from the literature resonates with the current finding. Research on moral development suggests that friendships and peer interactions offer important contexts for moral growth by providing opportunities for cooperation, reflection, and reciprocity among peers (Eisenberg et al., 2015). According to a systematic review on adolescent identity development and peer relationships (Ragelienė, 2016), positive relationships with a peer group—characterized by mutual respect and acceptance—facilitates identity development (e.g., achieving a sense of identity). Greater peer support may foster identity development both generally and specifically, by promoting moral growth and prosocial identity. Thus, it is possible that peer support may play a role in amplifying the protective effect of prosocial identity on reducing youth recidivism.

However, the interpretation of this finding should also be informed by the current state of knowledge on the (not yet clear) links among prosocial identity, peer influence, and delinquent behavior. Although the literature generally supports that a positive relationship with peers is beneficial for adolescent identity development, "empirical research in this area is extremely limited" and it remains ambiguous *what* and *how* other individual and environmental factors might influence both the quality of peer relationships and identity development (Ragelienė, 2016, p. 97).

For example, compared to other youth, justice-involved youth may have less stable relational identities or more doubts about their commitment to friendships (Klimstra et al., 2011). These doubts imply less stable perceived peer support and perhaps more tenuous influence on the development of a prosocial identity and behavior.

More importantly, justice-involved youth are relatively likely to be exposed to antisocial peer influence. Antisocial and prosocial peers are risk and protective factors, respectively, for recidivism. However, there are skeptics who view this relation as "one of the most notorious examples of a spurious link" in criminology (McGloin and Thomas, 2019, p.241), arguing that deviant youth may either be left with no choice but to associate with, or actively select, deviant peers. The mechanisms of peer impact are not clear.

This complexity is partially illustrated by findings from the current study: while a higher level of prosocial identity (i.e., moral ideal self) better protects against reoffending among those with higher perceived peer support, it is intriguing to note that when the level of prosocial identity is low, higher peer support instead predicts a higher likelihood of reoffending (i.e., lower peer support becomes protective). Considering that the two items used to measure peer support in the study are general and neutral ("I feel supported by my friends," "My friends stand by me during difficult times"), a potential explanation could be that youth with a low level of prosocial identity may tend to befriend peers with similar self-perceptions, attitudes, and values, who may express support for deviant values and behaviors, thus reinforcing delinquent behavior.

Another angle to read this crossover interaction is to consider the context in which the two self-report questions on peer support were answered. Youth were answering those questions at a time when they were just arrested and/or placed on probation, and most of the

youth were low- or moderate-risk. It is probable that youth who had a pretty solid (i.e., moderate or high level of) prosocial sense of self *and* felt like their peers support them, even though they "messed up" by involving in the juvenile justice system, were less likely to reoffend. On the other side, the crossover interaction suggests that youth who endorsed a prosocial sense of self but were rejected by their peers (i.e., perceiving low peer support) were more likely to reoffend. Some literature indicates that narcissism and an inflated sense of self are risk factors for recidivism among justice-involved youth (e.g., Barry et al., 2007; Schalkwijk et al., 2016). It is possible that the prosocial identity measures in the current study capture some variance in this subset of youth who perhaps had an inflated sense of prosocial self, as indicated by peer rejection. This interpretation, however, is speculative.

Overall, without additional information from participants about the nature of their peer associations, it is challenging to draw definitive conclusions. For this finding, I recommend focusing on the stronger part of the effect, where peer support enhances the protective effect of prosocial identity. Future research that incorporates more comprehensive measures of peer networks and delinquency is needed to help clarify these dynamics.

Lack of Moderation Evidence: The Roles of Identity Internalization, Self-Efficacy, and School Connectedness

Apart from the interaction between prosocial identity and peer support, neither regression nor MSM analyses detected significant interaction effects between prosocial identity and other proposed moderators. These moderators include the prominence of prosocial identity as measured by moral internalization, general self-efficacy as a proxy measure for the validation of prosocial identity, and school connectedness. Several reasons could potentially explain the observation of these null effects.

The study may not have had sufficient statistical power to detect an interaction effect that actually exists. Although power analysis estimates, conducted when defining the PA subsample, suggested that a sample size of 760 youths would have enough statistical power to detect a point-biserial correlation of 0.18 with a 12.2% recidivism rate at a significance level of 0.05 and 80% power, this sample size may still be inadequate for detecting subtle interaction effects that add complexity to the statistical models. In fact, the strength of associations between identity measures and the reoffending outcome found in this study is much smaller than what has been observed in the literature. The point-biserial correlation between other-oriented identity and recidivism in the study is -0.08, and in regressions, the odds ratio of other-oriented identity is 0.99 or 0.95. In addition, the main effect of the holistic prosocial identity on recidivism is not significant in the study. Some important limitations, such as the short follow-up period and low base rate of reoffending, may explain the small effect sizes and null findings. These factors require a much larger sample size to achieve sufficient power to detect both main and interaction effects, if they exist.

Certain measures of variables may not be the most precise representations of the concepts of interest within the conceptual model the study aims to test. In addition to what I have discussed earlier about measures of conventional identity, using general self-efficacy as a proxy for prosocial identity validation is not ideal either. It largely captures a youth's overall

confidence in handling difficulties but not their confidence in achieving a potential prosocial identity specifically. This is a limitation of the data, as inaccurate measurement can introduce noise that obscures true relations, including nuanced patterns like interaction effects.

Potential collinearity between two constructs—moral ideal self and moral internalization—may complicate the task of distinguishing their individual and interaction effects. In the PA subsample, a moderate correlation exists between moral ideal self and moral internalization ($\tau = 0.43$, $p<0.001$), suggesting that these two scales might not be the most suitable measures for assessing the content and prominence of prosocial identity when the goal is to explore the potential interaction effects between them.

Finally, the study sample may lack sufficient diversity. The PA subsample has an age cutoff of 17 years due to practical considerations. Compared to the larger sample of youth from three states, the PA subsample is younger and has a lower baseline risk of recidivism, both of which contribute to lower observed one-year recidivism rates. Since interaction effects often emerge more clearly across a wide range of conditions or values, these sample characteristics can limit the range of observed data, thereby making it challenging to detect interactions.

Of course, despite strong theoretical promise, it is also possible that some of the proposed interaction effects simply do not exist. As this study is the first attempt to empirically examine theoretical relations from the conceptual model, future research is needed to replicate the methodologies, verify these findings, and explore the complex dynamics between identity factors and behavioral outcomes.

Prosocial Identity Measures: Application from General Youth Populations to Justice-Involved Youth Contexts

The secondary aim of the study is to examine the generalizability of measures of prosocial identity that were developed with non-referred populations, to justice-involved youth. The study used two measures to assess participants' degree of prosocial identity: the Moral Ideal Self Scale (MISS, Hardy et al., 2014) and the Moral Internalization Scale (Aquino & Reed, 2002). Furthermore, the study introduced two concepts—other-oriented identity and conventional identity—that highlight different aspects of prosocial identity content and were measured by subscales from the MISS. I found these measures of prosocial identity appear generalizable among justice-involved youth, though improvement in precision for the measure of conventional identity is favorable, as discussed earlier.

Both the MISS and the Moral Internalization Scale were developed from and have typically been applied to non-referred youth recruited at schools or online. Compared to youth in this study, those non-referred groups were more likely to include fewer racially minority youth and have more even gender distributions. Despite clear demographic and presumed socioeconomic differences, I found that the MISS, Moral Internalization Scale, and the concepts of other-oriented and conventional identity derived from the MISS, exhibited satisfactory psychometric properties. They showed good internal consistency, appropriate factor structures, and reasonable patterns of divergent and convergent relations with other variables.

These findings pave the way for future research aimed at deepening our understanding of prosocial identity among justice-involved youth. Previous studies exploring the relation between prosocial identity and youths' delinquent behaviors have used various operationalizations of prosocial identity (single items, proxy measures, etc.). Ideally, further research using validated measures of prosocial identity (and supplemental measures of feared antisocial selves) will enhance comparison and consistency across studies. There is significant scope for exploration in this area, ranging from basic insights, such as the developmental patterns of prosocial identity and its features and variations by demographics, socioeconomic factors, and justice involvement, to a more comprehensive understanding of how prosocial identity contributes to positive development in this population.

Such insights currently are lacking but could be invaluable for targeting prosocial identity in interventions. For example, in the context of prosocial identity development, consensus has yet to be reached even within the general youth population. A recent perspective offered by Krettenauer (2022) suggests that there are no absolute starting or ending points in terms of age or development for moral identity (which, as previously reviewed, is often equated with prosocial identity in the developmental literature). Rather, moral identity may be viewed as a goal with characteristics that evolve systematically from concrete to abstract and from externally to internally motivated throughout development.

This perspective introduces greater flexibility for efforts aimed at fostering prosocial identity across different age groups—children, adolescents, and adults. However, its applicability to the justice-involved population needs further examination, especially because research indicates that youth involved in the juvenile justice system often demonstrate more instability in identity formation (e.g., diffused in identity status, less exploration or commitment to educational identity, higher level of reconsideration on interpersonal identity), compared to their peers in the general population—Klimstra et al., 2011.

Additionally, preliminary results from the current study suggest that prosocial identity measures vary slightly by gender, with girls scoring higher in moral ideal self. This aligns with literature findings that females tend to exhibit more prosocial tendencies than males in both moral (Eisenberg et al., 2015; Xu et al., 2023) and conventional (Rocque et al., 2016) self-concepts. Understanding these nuances can inform the development of tailored intervention programs for justice-involved youth, taking into account group differences.

It is important to mention again that both the MISS and the Moral Internalization Scale are validated tools operationalizing prosocial identity as moral identity, defined as the "extent to which individuals identify with, and are invested in, being a moral person and doing what is moral or right" (Hardy et al., 2014, p. 45). This operationalization is often emphasized in psychological research, leaning towards an other-oriented identity characterized by an intention and behavior to benefit others. It captures less of what is meant by a conventional identity, which primarily involves conforming to laws, pursuing basic educational and vocational goals, and taking typical adult roles and responsibilities—a focus that is often highlighted in criminological studies. Findings from studies utilizing these measures of prosocial identity should be interpreted with this perspective in mind. Furthermore, future research can refine

the conventional identity measure and add supplemental measures for feared possible selves, as mentioned earlier, to reflect the conceptualization of prosocial identity in the criminological literature and the propositions of the Jian & Skeem (2023) model.

Study Strengths and Limitations

This study uses the most recent data from a relatively large sample of justice-involved youth (*N*=760) who were referred to community supervision across five counties in Pennsylvania from July 2021 to February 2023. It follows their recidivism outcomes over a one-year period to evaluate some theoretical hypotheses from an intervention framework centered on prosocial identity (Jian & Skeem, 2023). The study applies a range of analytical techniques to examine these hypotheses. It is the first one, to the best of my knowledge, that uses validated measures of prosocial identity to assess its impact on recidivism among justice-involved youth. Findings from the study confirm that prosocial identity measures developed for the general youth population are appropriate for justice-involved youth. The study also clarifies and provides partial empirical support for theoretical hypotheses from the identity-based intervention framework.

The study has several limitations that are important to consider for interpreting the results and planning future research. *First,* the observational nature of the study is not optimal for testing the impact of prosocial identity on recidivism outcomes. An experimental design that manipulates some aspects of prosocial identity, adopting intervention strategies outlined by Jian & Skeem (2023), can better establish causality and understand the role of prosocial identity in preventing delinquent behavior.

Second, there are limitations within the data. As previously mentioned, recidivism records from the criminal justice system—for those who aged out to older than 18 during the one-year follow-up—were not available at the time of data analysis. To ensure complete recidivism records, I excluded youths who aged above 17 at the start of the recidivism tracking from analyses. This decision for the age cutoff was based on practical rather than theoretical considerations, which might limit the generalizability of the findings. Moreover, due to the ongoing nature of the larger project from which this study's data were sourced, the youths' placement data had yet to be processed at the time of this dissertation work. This means information on which recidivism offenses took place within institutions—data typically excluded from analyses—is missing. It would be ideal to replicate the current study when complete recidivism records and placement information are available.

Third, as previously addressed, there are limitations in the measures. The measure of conventional identity can be refined to align more closely with the conceptualization of prosocial identity in criminology; the measure of the feared possible self is lacking; and using self-efficacy as a proxy measure for prosocial identity validation is not ideal. Finally, there is potential collinearity between moral ideal self and moral internalization, which may result in unstable coefficient estimates in regression models where the two variables are included simultaneously.

These limitations highlight the need for further research that uses more rigorous designs, comprehensive data, and refined measures to thoroughly explore the impact of prosocial identity on reducing recidivism among justice-involved youth.

Conclusion and Recommendations

The Positive Youth Justice (PYJ) paradigm is gaining traction. Shifting from risk-reduction to strengths-focused approaches opens new avenues to help young people build prosocial assets and achieve overall positive development, including desistance from antisocial and delinquent behavior. Centering on prosocial identity as a promising target to begin translating the largely abstract PYJ philosophy into practice, this dissertation uses empirical data mostly drawn from a sample of 760 court-referred youths to examine components of a treatment-relevant theory proposed by Jian and Skeem (2023).

The most important findings from the study are that 1) unlike conventional identity, other-oriented identity predicts youth recidivism; and 2) peer support significantly moderates the relation between prosocial identity and recidivism. These findings provide an initial clarification on the definition of prosocial identity as a treatment target in the juvenile justice context, and partially confirm the hypothesis from the Jian & Skeem (2023) model that social support is crucial and can enhance the protective effects of prosocial identity against youth reoffending. The null effects of holistic prosocial identity (i.e., moral ideal self) and other proposed moderators (i.e., moral internalization, self-efficacy, caregiver support, school connectedness) on recidivism raise unsolved puzzle that necessitates future research in this area, to advance our understanding of the potential mechanism of change centering on prosocial identity and to inform intervention strategies in practice.

Future research may consider the following steps: 1) replicating and expanding upon the current study with more comprehensive and diverse data, including a wider range of ages, baseline risks, longer recidivism follow-ups, and other justice outcomes (e.g., probation outcomes, completion of therapeutic programs, recidivism after the completion of probation). This can refine and validate findings from the current study. Replication of the current study is possible once the complete recidivism and placement records of the multi-state sample are available later this year. 2) Developing and employing more accurate measures for conventional identity, feared self, and the notion of prosocial identity validation that capture youth's sense of confidence in achieving their possible prosocial self-images. Refining these measures can better test the propositions of the Jian & Skeem (2023) model, build a consensus on the definition and measurement of prosocial identity in the juvenile justice context, and thus facilitate research that delves into its developmental patterns, variations by demographics and socioeconomic factors, and its role in positive development. 3) Transitioning from observational to experimental study designs to manipulate aspects of prosocial identity, rigorously testing intervention strategies that foster a balanced prosocial possible self (i.e., pairing the hoped-for prosocial self with a feared self), enhance its prominence, and validate it both internally and externally. Such research will further illuminate the conditions under which prosocial identity can most effectively protect against reoffending and guide the design of real-world identity-based intervention programs for at-risk and justice-involved youth.

On the practical side, findings from the study may guide prevention and intervention efforts, including those from the legal systems, service providers, families, and the broader community, to focus more on youths' internal processes and their interactions with contextual factors such as social support from peers. The study suggests that juvenile justice interventions aiming to leverage prosocial identity to reduce reoffending should prioritize cultivating motivations and values that encourage a desire to benefit others, alongside promoting prosocial behaviors such as caring, helping, sharing, and contributing. This approach goes beyond simply pursuing educational or vocational goals or adhering to laws and social norms. Such interventions should also enhance peer support to achieve greater effects in curbing recidivism. School- and community-based programs that integrate identity shifts and social support may achieve synergistic effects in delinquency interventions.

References

Abrams, L. S., & Aguilar, J. P. (2005). Negative trends, possible selves, and behavior change: A qualitative study of juvenile offenders in residential treatment. *Qualitative Social Work, 4*(2), 175–196. https://doi.org/10.1177/ 1473325005052392

Ahadi, S., & Diener, E. (1989). Multiple determinants and effect size. *Journal of Personality and Social Psychology, 56*, 398-406.

American Psychological Association. (n.d.). Prosocial. In *APA dictionary of psychology*. https://dictionary.apa.org/prosocial

Andrews, D. A. (2011). The Risk-Need-Responsivity (RNR) Model of correctional assessment and treatment. In J. A. Dvoskin, J. L. Skeem, R. W. Novaco, & K. S. Douglas (Eds.), *Using social science to reduce violent offending*. Oxford Scholarship Online. https://doi.org/10.1093/acprof:oso/9780195384642.001.0001

Aquino, K. F., & Reed, A., II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423–1440. https:// doi.org/10.1037/0022-3514.83.6.1423

Baglivio, M. T., Jackowski, K., Greenwald, M. A., & Howell, J. C. (2014). Serious, violent, and chronic juvenile offenders. *Criminology & Public Policy, 13*(1), 83–116. https://doi.org/10.1111/1745-9133.12064

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Barnes-Lee, A. R., & Petkus, A. (2023). A scoping review of strengths-based risk and needs assessments for youth involved in the juvenile legal system. *Children and Youth Services Review*, *148*, Article 106878. https://doi.org/ 10.1016/j.childyouth.2023.106878

Barry, C. T., Grafeman, S. J., Adler, K. K., & Pickard, J. D. (2007). The relations among narcissism, self-esteem, and delinquency in a sample of at-risk adolescents. *Journal of Adolescence, 30*(6), 933–942. https://doi.org/10.1016/j.adolescence.2006.12.003

Berkman, E. T., Livingston, J. L., & Kahn, L. E. (2017a). Finding the "self" in self-regulation: The identity-value model. *Psychological Inquiry, 28*(2–3), 77–98. https://doi.org/10.1080/1047840X.2017.1323463

Berkman, E. T., Livingston, J. L., & Kahn, L. E. (2017b). The identity-value model of self-regulation: Integration, extension, and open questions. *Psychological Inquiry, 28*(2–3), 157–164. https://doi.org/10.1080/ 1047840X.2017.1343069

Brenner, P. S., Serpe, R. T., & Stryker, S. (2014). The causal ordering of prominence and salience in identity theory: An empirical examination. *Social Psychology Quarterly, 77*(3), 231–252. https://doi.org/10.1177/ 0190272513518337

Bryan, C. J., Adams, G. S., & Monin, B. (2013). When cheating would make you a cheater: Implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General, 142*(4), 1001–1005. https://doi.org/10 .1037/a0030655

Bryan, C. J., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences, 108*(31), 12653–12656. https://doi.org/10.1073/pnas.1103343108

Burke, P. (2020). Identity. In P. Kivisto (Ed.), *The Cambridge handbook of social theory* (1st ed., pp. 63–78). Cambridge University Press. https:// doi.org/10.1017/9781316677452.005

Burke, P. J., & Tully, J. C. (1977). The measurement of role identity. *Social Forces, 55*(4), 881–897. https://doi.org/10.2307/2577560

Bushway, S. D., & Paternoster, R. (2013). Desistance from crime: A review and ideas for moving forward. In C. L. Gibson & M. D. Krohn (Eds.), *Handbook of life-course criminology: Emerging trends and directions for future research* (pp. 213–231). Springer. https://doi.org/10.1007/ 978-1-4614-5113-6_13

Butler, D. C., Petterson, S., Phillips, R. L., & Bazemore, A. W (2013). Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery. *Health Services Research*, *48*(2), 539–559. https://doi.org/10.1111/j.1475-6773.2012.01449.

Butts, J. A. (2014). *Strengthening youth justice practices with developmental knowledge and principles.* National Institute of Corrections. https://nicic .gov/strengthening-youth-justice-practices-developmental-knowledge-and- principles

Butts, J. A., Bazemore, G., & Meroe, A. S. (2010). *Positive youth justice—Framing justice interventions using the concepts of positive youth development*. Coalition for Juvenile Justice.

Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *The ANNALS of the American Academy of Political and Social Science, 591*(1), 98–124. https://doi.org/10.1177/0002716203260102

Clark, L. F., Miller, K. S., Nagy, S. S., Avery, J., Roth, D. L., Liddon, N., & Mukherjee, S. (2005). Adult identity mentoring: Reducing sexual risk for African-American seventh grade students. *Journal of Adolescent Health, 37*(4), 337.e1–337.e10. https://doi.org/10.1016/j.jadohealth.2004.09.024

Clinkinbeard, S. S., & Murray, C. I. (2012). Perceived support, belonging, and possible selves strategies among incarcerated juvenile offenders. *Journal of Applied Social Psychology, 42*(5), 1218–1240. https:// doi.org/10.1111/j.1559-1816.2011.00884.x

Cohn, A., Maréchal, M. A., & Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *The Review of Economic Studies, 82*(4), 1289–1308. https://doi.org/10.1093/restud/rdv025

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience, 13*(9), 636–650. https://doi.org/10.1038/nrn3313

Cuevas, C., Wolff, K. T., & Baglivio, M. T. (2017). Self-efficacy, aspirations, and residential placement outcomes: Why belief in a prosocial self matters. *Journal of Criminal Justice, 52*, 1–11. https://doi.org/10.1016/j.jcrimjus .2017.06.006

Dahl, R. E., Allen, N. B., Wilbrecht, L., & Suleiman, A. B. (2018). Importance of investing in adolescence from a developmental science perspective. *Nature, 554*(7693), 441–450. https://doi.org/10.1038/nature25770

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. Perspectives on Psychological *Science, 14*(3), 481–496. https://doi.org/10.1177/1745691618804166

Eisenberg, N., Spinrad, T. L., & Knafo-Noam, A. (2015). Prosocial development. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Volume 3. Socioemotional processes* (7th ed., pp. 610–656). Wiley.

Erikson, E. H. (1968). *Identity: Youth and crisis*. Norton.

Flournoy, J. (2021, November 12). Why and how to care about covariates in longitudinal data [Conference presentation]. *The 7th annual UC Adolescence Consortium Institute*, CA, United States

Furlong, M. J., O'brennan, L. M., & You, S. (2011). Psychometric properties of the Add Health School Connectedness scale for 18 sociocultural groups. *Psychology in the Schools, 48*(10), 986–997. https://doi.org/10.1002/pits.20609

Giordano, P. C. (2022). Some cognitive transformations about the dynamics of desistance. *Criminology & Public Policy, 21*(4), 787–809. https://doi.org/10.1111/1745-9133.12609

Giordano, P. C., Cernkovich, S. A., & Rudolph, J. L. (2002). Gender, crime, and desistance: Toward a theory of cognitive transformation. *American Journal of Sociology, 107*(4), 990–1064. https://doi.org/10.1086/343191

Goodson, M. V., & Morash, M. (2017). Court-involved girls' perceptions of the attainability of a desired possible self and its connection to past adversity and current behavior. *Feminist Criminology, 12*(4), 384–404. https:// doi.org/10.1177/1557085116665189

Grether, T., JSowislo, J. F., & Wiese, B. S. (2018). Top-down or bottom-up? Prospective relations between general and domain-specific self-efficacy beliefs during a work-family transition. *Personality and Individual Differences, 121*, 131-139. https://doi.org/10.1016/j.paid.2017.09.021.

Haines, K., & Case, S. (2018). The future of youth justice. *Youth Justice, 18*(2), 131–148. https://doi.org/10.1177/1473225418791416

Hardy, S. A., Bean, D. S., & Olsen, J. A. (2015). Moral identity and adolescent prosocial and antisocial behaviors: Interactions with moral disengagement and self-regulation. *Journal of Youth and Adolescence, 44*(8), 1542– 1554. https://doi.org/10.1007/s10964-014-0172-1

Hardy, S. A., Francis, S. W., Zamboanga, B. L., Kim, S. Y., Anderson, S. G., & Forthun, L. F. (2013). The roles of identity formation and moral identity in college student mental health, health risk behaviors, and psychological well-being. *Journal of Clinical Psychology, 69,* 364–382. http://doi.org/10.1002/jclp.21913

Hardy, S. A., Walker, L. J., Olsen, J. A., Skalski, J. E., & Basinger, J. C. (2011). Adolescent naturalistic conceptions of moral maturity. *Social Development, 20*, 562–586. doi:10.1111/j.1467-9507.2010.00590.x

Hardy, S. A., Walker, L. J., Olsen, J. A., Woodbury, R. D., & Hickman, J. R. (2014). Moral identity as moral ideal self: Links to adolescent outcomes. *Developmental Psychology, 50*(1), 45–57. https://doi.org/10.1037/a0033598

Hendry, M. C., Douglas, K. S., Winter, E. A., & Edens, J. F. (2013). Construct measurement quality improves predictive accuracy in violence risk assessment: An illustration using the Personality Assessment Inventory. *Behavioral Sciences & the Law, 31*(4), 477–493. https://doi.org/10.1002/bsl.2066

Hoge, R. D., & Andrews, D. A. (2011). *Youth level of service/Case management inventory 2.0 (YLS/CMI 2.0): User's manual*. Multi-Health Systems. https://doi.org/10.1080/15283488.2021.1954519

Horowitz, E., Oyserman, D., Dehghani, M., & Sorensen, N. (2020). Do you need a roadmap or can someone give you directions: When school-focused possible identities change so do academic trajectories. *Journal of Adolescence*, *79*, 26–38. https://doi.org/10.1016/j.adolescence.2019.12.013

Jenkins, M., & Griffith, R. (2004). Using Personality Constructs to Predict Performance: Narrow or Broad Bandwidth. *Journal of Business and Psychology, 19*(2), 255–269. https://doi.org/10.1007/s10869-004-0551-9

Jian, L., & Skeem, J. L. (2023, December 7). Getting Traction on Positive Youth Justice: Prosocial Identity as a Promising Target for Intervention. *Psychology, Public Policy, and Law*. Advance online publication. https://dx.doi.org/10.1037/law0000410

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141–151. http://dx.doi.org/10.1177/001316446002000116

Kanacri, B. P. L., Pastorelli, C., Eisenberg, N., Zuffianò, A., & Caprara, G. V. (2013). The development of prosociality from adolescence to early adult- hood: The role of effortful control. *Journal of Personality*, 81(3), 302–312. https://doi.org/10.1111/jopy.12001

Kavussanu, M., Stanger, N., & Ring, C. (2015). The effects of moral identity on moral emotion and antisocial behavior in sport. *Sport, Exercise, and Performance Psychology, 4*(4), 268–279. https://doi.org/10.1037/ spy0000040

Killen, M., & Smetana, J. G. (2015). Origins and Development of Morality. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Volume 3. Socioemotional processes* (7th ed., pp. 701–749). Wiley.

Klimstra, T. A., Crocetti, E., Hale, W. W., Kolman, A. I.M., Fortanier, E., & Meeus, W. H. J. (2011). Identity formation in juvenile delinquents and clinically referred youth. *European Review of Applied Psychology / Revue Européenne de Psychologie Appliquée, 61*(3), 123–130. https://doi.org/10.1016/j.erap.2011.05.002

Krettenauer, T. (2011). The dual moral self: Moral centrality and internal moral motivation. *The Journal of Genetic Psychology: Research and Theory on Human Development, 172*(4), 309–328. https://doi.org/10.1080/00221325.2010.538451

Krettenauer, T. (2022). Development of moral identity: From the age of responsibility to adult maturity. *Developmental Review, 65*. https://doi.org/10.1016/j.dr.2022.101036.

Laub, J. H., & Sampson, R. J. (2003). *Shared beginnings, divergent lives: Delinquent boys to age 70.* Harvard University Press.

LeBel, T. P., Burnett, R., Maruna, S., & Bushway, S. (2008). The "Chicken and Egg" of subjective and social factors in desistance from crime. *European Journal of Criminology, 5*(2), 131–159. https://doi.org/10.1177/ 1477370807087640

Lewin, K. (1943). Psychology and the process of group living. *Journal of Social Psychology, 17*(1), 113–131. https://doi.org/10.1080/00224545 .1943.9712269

Liebenberg, L., Ungar, M., & Van de Vijver, F. (2012). Validation of the Child and Youth Resilience Measure-28 (CYRM-28) among Canadian youth. *Research on Social Work Practice, 22*(2), 219–226. https://doi.org/10.1177/1049731511428619

Lönnfjord, V., & Hagquist, C. (2018). The psychometric properties of the swedish version of the general self-efficacy scale: A Rasch analysis based on adolescent data. *Current Psychology, 37*, 703–715.

Lorah, J. A. (2020). Interpretation of main effects in the presence of non-significant interaction effects. *The Quantitative Methods for Psychology, 16*(1), 33–45. https://doi.org/10.20982/tqmp.16.1.p033

Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine, 43*(2), S8–S23. https://doi.org/10.1016/j.amepre.2012 .04.029

Löve, J., Moore, C. D., & Hensing, G. (2012). Validation of the Swedish translation of the general self-efficacy scale. *Quality of Life Research, 21*(7), 1249–1253.

Maitner, A. T., Mackie, D. M., Claypool, H. M., & Crisp, R. J. (2010). Identity salience moderates processing of group-relevant information. *Journal of Experimental Social Psychology, 46*(2), 441–444. https:// doi.org/10.1016/j.jesp.2009.11.010

Marcia, J. E. (1980). Ego identity development. In J. Adelson (Ed.), *Handbook of adolescent psychology* (pp. 159–187). Wiley.

Marin, L., Ruiz, S., & Rubio, A. (2009). The role of identity salience in the effects of corporate social responsibility on consumer behavior. *Journal of Business Ethics, 84*(1), 65–78. https://doi.org/10.1007/s10551-008-9673-8

Maruna, S. (2001). *Making good: How ex-convicts reform and rebuild their lives*. American Psychological Association.

McCall, G. J., & Simmons, J. L. (1978). *Identities and interactions: An examination of human associations in everyday life*. Free Press.

McGloin, J. M., & Thomas, K. J. (2019). Peer influence and delinquency. *Annual Review of Criminology, 2*, 241–262. https://doi.org/10.1146/annurev-criminol-011518-024551

McLean, K. C., Syed, M., & Shucard, H. (2019, April 22). *Bringing identity content to the fore: Links to identity development processes.* https:// doi.org/10.31234/osf.io/7jtxv

Meehl, P. E. (1971). High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology, 77*(2), 143–148. https://doi.org/10.1037/h0030750

Murphy, M. C., Taylor, V. J. (2012). The role of situational cues in signaling and maintaining stereotype threat. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and applications* (pp. 17–33). Oxford University Press.

Na, C., & Paternoster, R. (2019). Prosocial identities and youth violence. *Journal of Research in Crime and Delinquency, 56*(1), 84–128. https:// doi.org/10.1177/0022427818796552

Neugebauer, R., & van der Laan, M.J. (2007). Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference, 137*(2)*,* 419–434.

Odgers, C. L., Allen, N. B., Pfeifer, J. H., Dahl, R. E., Nesi, J., Schueller, S. M., Williams, J. L., & the National Scientific Council on Adolescence. (2022). *Engaging, safe, and evidence-based: What science tells us about how to promote positive development and decrease risk in*

online spaces (Council Report No. 2). UCLA Center for the Developing Adolescent. https://doi.org/10.31234/osf.io/rvn8q

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: race and the disciplining of young students. *Psychological science, 26*(5), 617–624. https://doi.org/10.1177/0956797615570365

Orbis Partners. (2000). *Youth Assessment Screening Inventory (YASI).* Orbis Partners.

Oyserman, D. (2007). Social identity and self-regulation. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 432–453). Guilford.

Oyserman, D., & Destin, M. (2010). Identity-based motivation: Implications for intervention. *The Counseling Psychologist, 38*(7), 1001–1043. https:// doi.org/10.1177/0011000010374775

Oyserman, D., & Markus, H. (1990). Possible selves in balance: Implications for delinquency. *Journal of Social Issues, 46*(2), 141–157. https://doi.org/ 10.1111/j.1540-4560.1990.tb01927.x

Oyserman, D., & Saltz, E. (1993). Competence, delinquency, and attempts to attain possible selves. *Journal of Personality and Social Psychology, 65*(2), 360–374. https://doi.org/10.1037/0022-3514.65.2.360

Oyserman, D., Bybee, D., & Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology, 91*(1), 188–204. https://doi.org/10 .1037/0022-3514.91.1.188

Oyserman, D., Elmore, K., & Smith, G. (2012). Self, self-concept, and iden- tity. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd ed., pp. 69–104). The Guilford Press.

Oyserman, D., Fryberg, S. A., & Yoder, N. (2007). Identity-based motivation and health. *Journal of Personality and Social Psychology, 93*(6), 1011– 1027. https://doi.org/10.1037/0022-3514.93.6.1011

Oyserman, D., Terry, K., & Bybee, D. (2002). A possible selves intervention to enhance school involvement. *Journal of Adolescence, 25*(3), 313–326. https://doi.org/10.1006/jado.2002.0474

Paternoster, R., & Bushway, S. (2008). Desistance and the feared self: Toward an identity theory of criminal desistance criminology. *Journal of Criminal Law and Criminology, 99*(4), 1103–1156.

Paternoster, R., Bachman, R., Kerrison, E., O'Connell, D., & Smith, L. (2016). Desistance from crime and identity: An empirical test with survival time. *Criminal Justice and Behavior, 43*(9), 1204–1224. https://doi.org/10.1177/0093854816651905

Pearl, J. (2010). An introduction to causal inference. *The International Journal of Biostatistics, 6*(2). http://doi.org/10.2202/1557-4679.1203

Petersen, M.L., & van der Laan, M. (2014). Causal models and learning from data: Integrating causal modeling and statistical estimation. *Epidemiology, 25*, 418–426. http://doi.org/ 10.1097/EDE.0000000000000078

Pfattheicher, S., Nielsen, Y. A., & Thielmann, I. (2022). Prosocial behavior and altruism: A review of concepts and definitions. *Current Opinion in Psychology, 44*, 124–129. https://doi.org/10.1016/j.copsyc.2021.08.021

Pfeifer, J. H., & Berkman, E. T. (2018). The development of self and identity in adolescence: Neural evidence and implications for a value-based choice perspective on motivated behavior. *Child Development Perspectives, 12*(3), 158–164. https://doi.org/10.1111/cdep.12279

Pfeifer, J. H., Lieberman, M. D., & Dapretto, M. (2007). "I know you are but what am I?!": Neural bases of self- and social knowledge retrieval in children and adults. *Journal of Cognitive Neuroscience, 19*(8), 1323–1337. https://doi.org/10.1162/jocn.2007.19.8.1323

Pfeifer, J. H., Masten, C. L., Borofsky, L. A., Dapretto, M., Fuligni, A. J., & Lieberman, M. D. (2009). Neural correlates of direct and reflected self- appraisals in adolescents and adults: When social perspective-taking informs self-perception. *Child Development, 80*(4), 1016–1038. https:// doi.org/10.1111/j.1467-8624.2009.01314.x

Phillips, T. M., & Pittman, J. F. (2003). Identity processes in poor adolescents: Exploring the linkages between economic disadvantage and the primary task of adolescence. *Identity: An International Journal of Theory and Research, 3*(2), 115–129. https://doi.org/10.1207/S1532706XID030202

Piquero, A., Farrington, D., & Blumstein, A. (2003). The criminal career paradigm: Background and recent developments. *Crime and Justice, 30*, 359–506.

Polley, E., LeDell, E., Kennedy C., Lendle, S., & van der Laan, M. (2021). *SuperLearner: super learner prediction.* R package, version 2.0–28. https://CRAN.R-project.org/package=SuperLearner

Pratt, M. W., Hunsberger, B., Pancer, S. M., & Alisat, S. (2003). A longitudinal analysis of personal values socialization: Correlates of a moral self-ideal in late adolescence. *Social Development, 12*(4), 563–585. https://doi.org/10.1111/1467-9507.00249

Ragelienė T. (2016). Links of adolescents identity development and relationship with peers: A systematic literature review. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie canadienne de psychiatrie de l'enfant et de l'adolescent*, *25*(2), 97–105.

Research & Evaluation Center. (2023, July). *Positive youth justice—Overview*. John Jay College of Criminal Justice. https://johnjayrec.nyc/pyj/

Rivnyák, A., Péley, B., Zsidó, A. N., & Láng, A. (2022). The impact of socioeconomic status upon adolescent identity formation: The mediating role of family characteristics. *An International Journal of Theory and Research, 22*(4), 337–346.

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology (Cambridge, Mass.)*, *11*(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

Rocque, M., Posick, C., & Paternoster, R. (2016). Identities through time: An exploration of identity change as a cause of desistance. *Justice Quarterly, 33*(1), 45–72. https://doi.org/10.1080/07418825.2014.894111

Sampson, R. J., & Laub, J. H. (1997). A life course theory of cumulative dis- advantage and the stability of delinquency. In T. P. Thornberry (Ed.), *Advances in criminological theory: Volume 7. Developmental theories of crime and delinquency* (pp. 133–161). Transaction Publishers.

Saulnier, L., & Krettenauer, T. (2023). Internet impropriety: Moral identity, moral disengagement, and antisocial online behavior within an early adolescent to young adult sample. *Journal of Adolescence, 95*(2), 264–283. https://doi.org/10.1002/jad.12112

Schalkwijk, F.W., Dekker, J., Peen, J., Stams, G. J. (2016). Narcissism, self-esteem, shame regulation and juvenile delinquency. *Forensic Science and Criminology*, *1*(1), 1–5. https://doi.org/10.15761/FSC.1000105

Scholz, U., Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*(3), 242–251.

Schwab, J., Lendle, S., Petersen, M., van der Laan, M., & Gruber, S. (2020). *Ltmle: longitudinal targeted maximum likelihood estimation.* R package, version 1.2–0. https://CRAN.R-project.org/package=ltmle

Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). NFER-NELSON.

Scott, E., & Steinberg, L. (2008). *Rethinking juvenile justice*. Harvard University Press.

Shadish, W. R. (1995). The logic of generalization: Five principles common to experiments and ethnographies. *American Journal of Community Psychology, 23*(3), 419–428. https://doi.org/10.1007/BF025 06951

Skeem, J., & Kennealy, P. (2015, March). Role of strengths in protecting high-risk youth against violent infractions and re-arrest. In R. Morgan (Chair), *Working to improve correctional practice: Assessing risk, community intervention, and understanding desistance*. Symposium presented at the annual conference of the American Psychology-Law Society, San Diego, CA, United States.

Social Deprivation Index (SDI). (2021, November 29). *Robert Graham Center - Policy Studies in Family Medicine & Primary Care.* https://www.graham-center.org/rgc/maps-data-tools/sdi/social-deprivation-index.html.

Steele, C. M. (2010). *Whistling Vivaldi: How stereotypes affect us and what to do*. W.W. Norton.

Stryker, S. (1987). Identity theory: Developments and extensions. In K. Yardley & T. Honess (Eds.), *Self and identity: Psychosocial perspectives* (pp. 89–103). Wiley.

Stryker, S., & Serpe, R. T. (1994). Identity salience and psychological centrality: Equivalent, overlapping, or complementary concepts? *Social Psychology Quarterly, 57*(1), 16–35. https://doi.org/10.2307/ 2786972

Ungar, M., & Liebenberg, L. (2012). The Child and Youth Resilience Measure-28: User manual. Halifax, NS: Resilience Research Centre, Dalhousie University.

van der Laan M. J. & Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer.

Van Vugt, E., Gibbs, J., Stams, G. J., Bijleveld, C., Hendriks, J., & van der Laan, P. (2011). Moral development and recidivism: A meta-analysis. *International journal of offender therapy and comparative criminology*, *55*(8), 1234–1250. https://doi.org/10.1177/0306624X11396441

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*(6023), 1447–1451. https://doi.org/10.1126/science.1198364

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review, 125*(5), 617–655. https://doi.org/10.1037/rev0000115

Williams, J. L., Mims, L. C., & Johnson, H. E. (2019). *Young adolescent development*. University of Virginia Curry School of Education & Human Development. https://education.virginia.edu/documents/ynworking-paper- young-adolescent-development01-2019pdf

Xu, X., Li, M., Kwan, H. K., & Zhang, X. (2023, June 3). The antecedents of moral identity: A meta-analytic review. Asia Pacific Journal of Management. Advance online publication. https://doi.org/10.1007/s10490-023-09891-8

Yeager, D. S., Dahl, R. E., & Dweck, C. S. (2018). Why interventions to influence adolescent behavior often fail but could succeed. *Perspectives on Psychological Science, 13*(1), 101–122. https://doi.org/10.1177/ 1745691617722620

Zhang, M., Yu, Y., Wang, S., Salvatore, M., Fritsche, L. G., He, Z., Mukherjee, B. (2020). Interaction analysis under misspecification of main effects: Some common mistakes and simple solutions. *Statistics in Medicine*, *39*, 1675–1694. https://doi.org/10.1002/sim.8505

Appendix A: Baseline Measures

Moral Ideal Self Scale (Hardy et al., 2014)

e.g., How much does this describe the type of person you really want to be? "generous"
(1- Not at all | 2 - | 3 - | 4 - Neutral | 5 - | 6 - | 7 - Very much)

- Loving/Caring (9 items): generous, understanding, thankful, compassionate, loving, forgiving, considerate, caring, helpful.
- Honest/True (3 items): truthful, true, loyal
- Integrity (2 items): stands up for his/her beliefs, follows values
- Virtuous (2 items): respectful, responsible
- Knows/Choose Right (4 items): good example, makes good choices, does good actions, has good values

Other-oriented identity is measured by items in the Loving/Caring parcel.
Conventional identity is measured by items in the Honest/True, Integrity, Virtuous, and Knows/Choose Right parcels.
Prosocial Identity in research questions 2, 3 and 4 is measured by all the 20 items in the Moral Ideal Self Scale.

Moral Internalization Scale (Aquino & Reed, 2002)

Listed below are some traits that may describe a person: caring, compassionate, fair, friendly, generous, helpful, hardworking, honest, kind. The person with these traits could be you or it could be someone else. For a moment, visualize in your mind the kind of person who has these traits. Imagine how that person would think, feel and act. When you have a clear image of what this person would be like, answer the following questions.
(1- Strongly Disagree | 2 - Disagree | 3 - Neutral| 4 - Agree | 5 – Strongly Agree)

1. *It would make me feel good to be a person who has these traits.*
2. *Being someone who has these traits is an important part of who I am.*
3. *I would be ashamed to be a person who has these traits.*
4. *Having these traits is not really important to me*
5. *I strongly desire or want to have these traits.*

Generalized Self-efficacy Scale (Schwarzer & Jerusalem, 1995)

(1 - Not at all true | 2 – Hardly true | 3 – Moderately true | 4 - Exactly true)

1. *I can always manage to solve difficult problems if I try hard enough*
2. *If someone opposes me, I can find the means and ways to get what I want.*
3. *It is easy for me to stick to my aims and accomplish my goals.*
4. *I am confident that I could deal efficiently with unexpected events.*
5. *Thanks to my resourcefulness, I know how to handle unforeseen situations.*
6. *I can solve most problems if I invest the necessary effort.*
7. *I can remain calm when facing difficulties because I can rely on my coping abilities.*
8. *When I am confronted with a problem, I can usually find several solutions.*
9. *If I am in trouble, I can usually think of a solution.*
10. *I can usually handle whatever comes my way.*

Social support – peer (items from CYRM-28, Ungar & Liebenberg, 2012)

(1 - Not at all | 2 - A little | 3 - Somewhat | 4 - Quite a bit | 5 - A lot)

1. *I feel supported by my friends.*
2. *My friends stand by me during difficult times.*

Social support – caregiver (items from CYRM-28, Ungar & Liebenberg, 2012)

(1- Not at all | 2 - A little | 3 - Somewhat | 4 - Quite a bit | 5 - A lot)

3. *My caregivers watch me closely.*
4. *My caregivers know a lot about me.*
5. *I eat enough most days.*
6. *I talk to my caregivers about how I feel.*
7. *My caregivers stand by me during difficult times.*
8. *I feel safe when I am with my caregivers.*
9. *I enjoy my caregivers' cultural and family traditions.*

School Connectedness Scale (Furlong et al., 2011)

(1- Strongly Disagree | 2 - Disagree | 3 - Neutral| 4 - Agree | 5 – Strongly Agree)

1. *I feel close to people at this school*
2. *I am happy to be at this school.*
3. *I feel like I am part of this school.*
4. *The teachers at this school treat students fairly.*
5. *I feel safe in my school.*

Brief Self Control (Tangney et al., 2004)

(1 - Not at all like me | 2 - Unlike me | 3 - Sometimes like me | 4 - Like me | 5 - Very much like me)

1. *I am good at resisting temptation.*
2. *I have a hard time breaking bad habits.*
3. *I am lazy.*
4. *I say inappropriate things*
5. *I do certain things that are bad for me, if they are fun.*
6. *I refuse things that are bad for me, even if they are fun.*
7. *I wish I had more self-discipline or self-control.*
8. *People would say that I have iron self-discipline or amazing self-control.*
9. *Pleasure and fun sometimes keep me from getting work done.*
10. *I get distracted easily.*
11. *I do things that feel good in the moment but regret later on.*
12. *Sometimes I can't stop myself from doing something, even if I know it is wrong.*
13. *I often act without thinking through all the alternatives.*

Youth Social Responsibility Scale (Pancer et al., 2007)

(-4 Very strongly disagree, 0 Neither agree nor disagree, +4 Very strongly agree)

1. *Helping others gives a person a tremendous feeling of accomplishment.*
2. *Young people have an important role to play in making the world a better place.*
3. *It is important for people to know what's going on in their communities.*
4. *People in their teens can't vote, so there is not really any reason for them to care about politics and government.*
5. *Everybody should volunteer some time for the good of their community.*
6. *It's important for people in their teens to know what's going on in the world.*
7. *By helping others, parents set an important example for their children.*
8. *Teenagers should just enjoy themselves and not worry about things like poverty and the environment.*
9. *Schools should stick to the basics and not spend so much time trying to teach students about moral or social issues.*
10. *People should help one another without expecting to get paid or rewarded for it.*

Note: Social Responsibility in the study was constructed by dropping items 4, 8, and 9 as these items had very low factor loadings to a single “social responsibility” factor.

Appendix B: Results from Psychometric Analyses of Baseline Measures

Table 1. Characteristics of the larger sample (*N*= 1,913)

	n (valid %)	Range	missing
Age (mean/SD)	15.6 (1.7)	9.2-20.9	36 (1.9%)
Age by Category			36 (1.9%)
<=14	618 (32.9%)		
15-17	1,147 (61.1%)		
>=18	112 (6.0%)		
Gender			9 (0.5%)
Male	1,379 (72.4%)		
Female	525 (27.6%)		
Race			119 (6.2%)
White	697 (38.8%)		
Black	864 (48.2%)		
Latinx	233 (13.0%)		
Risk Level[4]			132 (6.9%)
High	220 (12.4%)		
Moderate	770 (43.2%)		
Low	791 (44.4%)		

[4] The risk-need assessment used in PA was Youth Level of Service/Case Management Inventory Version 2.0 (YLS/CMI) (Hoge & Andrews, 2011), and in VA and WI were Youth Assessment and Screening Instrument (YASI; Orbis Partners, Inc., 2008). Here risk level for each youth in the larger sample takes values “Low”, “Moderate”, or “High” by combining risk assessment results from YLS and YASI.

Table 2. Bivariate correlations among constructs, using Kendall's rank-based correlation tests for ordinal variables without normal distribution (*N*=1,913). Testing statistics is Kendall's tau (τ)

	Other-oriented Identity	Conventional Identity	Internalization	Self-efficacy	Peer Support	Caregiver Support	School Connectedness	Self-control	Social Responsibility	Risk Level
Other-oriented Identity	1									
Conventional Identity	0.59***	1								
Prominence/ Internalization	0.40***	0.36***	1							
Validation/ Self-efficacy	0.23***	0.26***	0.16***	1						
Social Support - Peer	0.19***	0.19***	0.14***	0.22***	1					
Social Support - Caregiver	0.32***	0.31***	0.26***	0.27***	0.23***	1				
School Connectedness	0.21***	0.20***	0.18***	0.19***	0.21***	0.28***	1			
Self-control	0.13***	0.19***	0.10***	0.16***	0.08***	0.15***	0.09***	1		
Social Responsibility	0.40***	0.33***	0.40***	0.20***	0.15***	0.28***	0.20***	0.11***	1	
Risk level	-0.16***	-0.16***	-0.17***	-0.09***	-0.10***	-0.19***	-0.20***	-0.13***	-0.14***	1

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

Table 3. Description of baseline measures (*N*=1,913)

Measurement	Construct	Min.	Mean	Max.	SD	Reliability (Cronbach's α)
Moral Ideal Self Scale	Loving/Caring	1	5.93	7	0.95	0.89
	Honest/True	1	6.00	7	1.01	0.72
	Integrity	1	5.80	7	1.09	0.50
	Virtuous	1	5.98	7	1.09	0.66
	Knows/Choose Right	1	5.58	7	1.16	0.86
Moral Ideal Self Scale	Other-oriented Identity	1	5.93	7	0.95	0.89
	Conventional Identity	1	5.81	7	0.96	0.91
Moral Ideal Self Scale	Holistic Prosocial Identity	1	5.86	7	0.90	0.94
Moral Internalization Scale	Prosocial Identity Prominence	1	3.93	5	0.74	0.79
Generalized Self-efficacy Scale	Prosocial Identity Validation	10	29.94	40	4.67	0.81
Child and Youth Resilience Measure (CYRM-28)	Social Support - Peer	1	3.89	5	1.10	0.85
	Social Support - Caregiver	1	4.11	5	0.80	0.86
School Connectedness Scale	School Connectedness	5	17.15	25	4.54	0.84
Youth Social Responsibility Scale	Social responsibility	-4	2.26	4	1.27	0.85

Table 4. Comparing means scores of prosocial identity, self-efficacy, social support, and school connectedness measures from the multi-state sample to norm data on the general youth (*N*=1,913)

	Mean/SD in the multi-state sample	Mean/SD from the general youth	Difference in mean[1]	95% CI
Moral Ideal Self Scale	5.86 (0.90)	6.26[2] (0.86) 6.23[3] (0.90)	-0.40***	-0.49, -0.31
Other-oriented Identity	5.93 (0.95)			
Conventional Identity	5.81 (0.96)			
Moral Internalization	5.40[4](1.10)	6.10[5] (0.94)	-0.70***	-0.82, -0.58
Self-efficacy	29.94 (4.67)	29.55 (5.32)[7]		
Child and Youth Resilience Measure (CYRM-28) – Peer Support	3.87[6] (1.10)	4.13[8]		
Child and Youth Resilience Measure (CYRM-28) – Caregiver Support	4.11[6] (0.80)	4.22[8] (physical) 4.02[8] (psychological)		
School Connectedness	17.15 (4.54)	16.6 (4.3)[9]		

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

Note 1. Online Independent samples *t* test was conducted.

Note 2. Hardy et al., 2014, a local school adolescents' sample (*N*=510; age 10–18 years). The paper reported mean and SD at parcel level, thus the statistics here were the average of five parcels.

Note 3. Hardy et al., 2014, a national online adolescents' sample (*N*=383; 15–18 years of age). The paper reported mean and SD at parcel level, thus the statistics here were the average of five parcels.

Note 4. Scores were converted from 5-point to 7-point to be comparable with the Hardy et al (2015) study.

Note 5. Hardy et al., 2015., an online sample in the United States (*N* = 384; ages 15–18 years).

Note 6. Results from the current study are not comparable to the norm data because measures for peer and caregiver support in the study were constructed from CYRM-28 items.

Note 7. Scholz et al., 2002

Note 8. Ungar & Liebenberg, 2012

Note 9. Furlong et al., 2

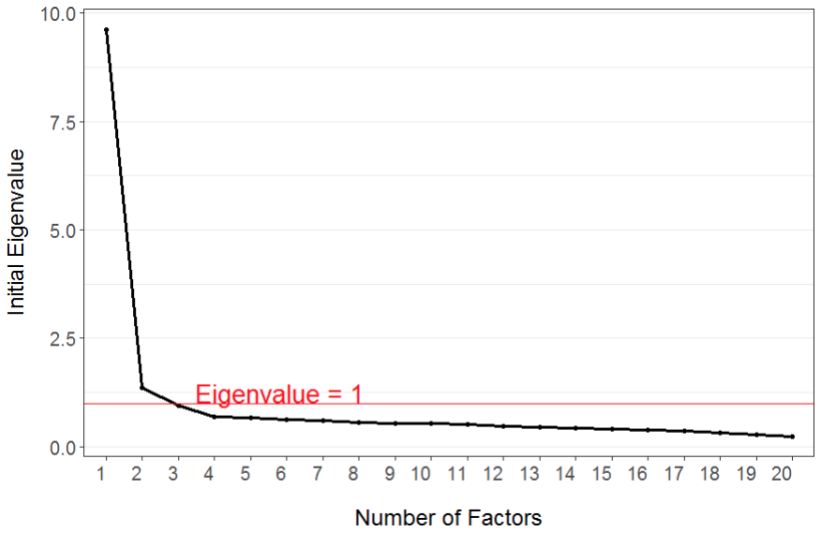

Figure 1. Optimal number of factors from exploratory factor analysis on the Moral Ideal Self Scale

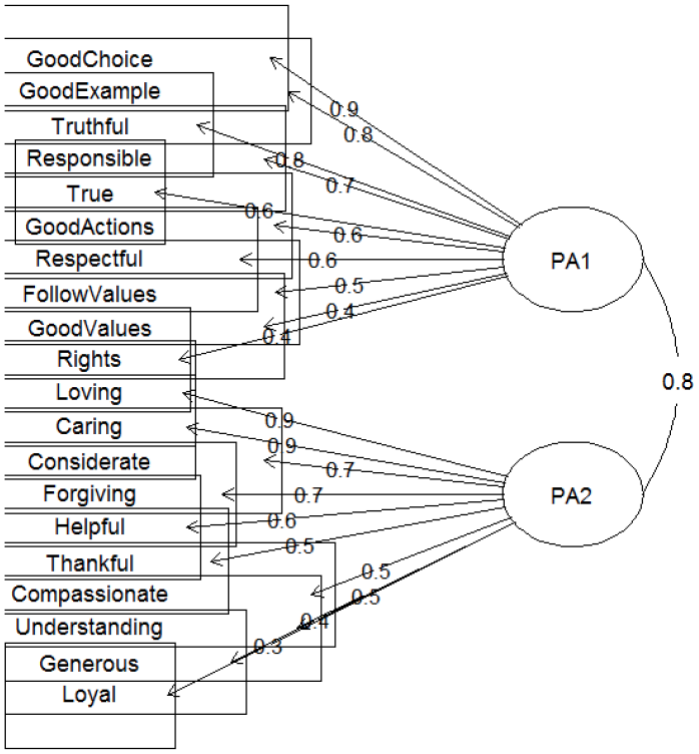

Figure 2. Factor loading and distribution of items when the number of factors is two, Rotate = “promax” (used oblique rotation to allow correlation between the two factors)

Table 5. Confirmatory Factor Analysis for prosocial identity measures, using item-level data from the Moral Ideal Self Scale (*N* = 1,912)

Model	χ^2	*p*	*df*	CFI	TLI	RMSEA
Model 1. Correlated five-factor analysis (the original Moral Ideal Self Scale)	1561.63	< 0.001	160	0.93	0.92	0.068
Model 2. Correlated two-factor analysis (other-oriented vs. conventional identity)	1746.80	< 0.001	169	0.92	0.91	0.070
Model 3. One factor analysis (other-oriented identity)	274.18	< 0.001	27	0.97	0.96	0.069
Model 4. One factor analysis (conventional identity)	679.18	< 0.001	44	0.94	0.92	0.870

Table 6. Confirmatory Factor Analysis for other baseline measures constructed from validated scales (*N* <=1,913, depending on number of missing values in items)

Model	χ^2	*p*	*df*	CFI	TLI	RMSEA
Model 1. Social support-caregiver (from Child and Youth Resilience Measure, CYRM-28, items 3,4,5,6,7,8,9) *N*=1,913	194.71	< 0.001	14	0.97	0.95	0.082
Model 2. Social responsibility (from Youth Social Responsibility Scale, items 1,2,3,5,6,7,10) *N*=1,895	155.79	< 0.001	14	0.97	0.96	0.073

Appendix C: Analytical Approaches to Address the Four Research Questions

A Directed Acyclic Graph Representing Background knowledge on Variables

Following the roadmap of causal modeling, I use a Directed Acyclic Graph (DAG, Pearl, 2010) to illustrate the relations among variables in the study. The DAG draws on knowledge from previous research and the conceptual model introduced in Figure 1.

Let A be the set of conditions of interest (or "exposure" in causal terminology), which, in this study, includes different *levels* of a set of main predictors such as prosocial identity content (operationalized as moral ideal self), prominence (operationalized as moral internalization), and validation (approximated by self-efficacy), along with social support (from peers and/or caregivers), and school connectedness. A_k represents the *k*th time-period when A is measured.

Let Y be a binary outcome, which is youth's reoffending behavior in the one-year follow-up measured by any new petition.

Let L_k be a set of covariates at the *k*th time, including static covariates such as race and gender and time-varying covariates such as age, socioeconomic status, and baseline risk. L_k is also confounders between A and Y.

Let U_k be unmeasured confounders that impact A, L, and Y. Variables in U_k are social, cultural, and contextual factors including norms, values, personal experience, life events, and so on.

In summary, we have endogenous variables L, A, and Y, and exogenous variables U, where:

- L is a set of baseline covariates consisting of
  - Age, continuous
  - Race, categorical (White, Latino, Black, other/unsure)
  - Gender, binary (male, female)
  - Socioeconomic status (SDI scores, continuous)
  - Baseline Risk (total risk scores, continuous)
- A is a set of conditions of interest (or "exposure")
  - In research question 1, A is prosocial identity content featured by other-orientation or conventionality, measured by scores from the loving/caring parcel and the other four parcels, respectively, in the Moral Ideal Self Scale. They are two continuous variables.
  - In research questions 2, 3, and 4, A is a set of variables include:
    - Holistic prosocial identity measured by scores on all items in the Moral Ideal Self Scale, continuous or categorical (low/high, or low/moderate/high)
    - Prosocial identity prominence (marker = moral internalization), continuous or categorical (low/high, or low/moderate/high)
    - Prosocial identity validation (marker = self-efficacy), continuous or categorical (low/high, or low/moderate/high)
    - Perception of social support from peer and/or caregiver, continuous or categorical (low/high, or low/moderate/high)
    - Perception of school connectedness, continuous or categorical (low/high, or low/moderate/high)

- Y is the outcome of interest
    - One-year recidivism, measured by any petition, binary.
- U is a set of unmeasured confounders
    - Social, cultural, and contextual factors including norms, values, personal experience, life events, etc.

Figure 1 presents a DAG that depicts relations among variables in a longitudinal view, as youth's identity is in development and their situations and behaviors may change over time. The focus is on the main effects of each variable in set A on the outcome Y, along with the potential interaction effects among variables in A (e.g., identity content X identity internalization, identity content X self-efficacy, identity content X social support, identity content X school connectedness) on the outcome Y. To denote an interaction term, represented as effect modifier V (as illustrated later in Figure 7) in conjunction with A in a DAG, A is substituted with A x V following the approach suggested by Flournoy (2021).

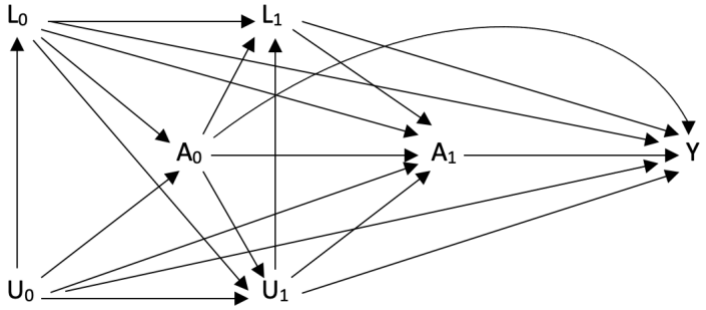

Figure 1. Relations among variables in a longitudinal view

The current study measures covariates L and conditions A only at baseline and follows recidivism records for one year. Thus, the longitudinal view in Figure 1 can be simplified to approximate a point-treatment scenario where the effect of A on Y (and the interaction effect of A and V on Y) within one time period is investigated. Relations among variables are depicted in Figure 2.

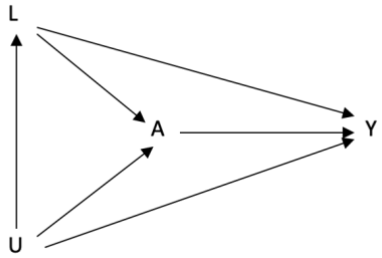

Figure 2. Relations among variables within one time period

Based on Figure 2, identifying causal relations between A and Y and between the interaction of A and V on Y needs to make assumptions on the unmeasured confounders U. A convenient assumption is that U does not exist (i.e., assuming no unmeasured confounding, see Figure 3). Alternatively, we can assume that all the unmeasured factors are independent (i.e.,

having separate U_A, U_L, and U_Y, see Figure 4), or either A or Y is not impacted by U (see Figure 5 and Figure 6). According to back-door criteria for causal identification (Pearl, 2010), under any of these assumptions, if we adjust for measured confounders L and if we have sufficient support from data (e.g., rare positivity violations, which means almost each condition of A occurs for each possible combination of strata within covariates L), the causal relations can be identified.

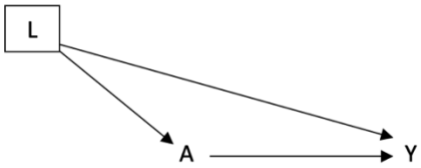

Figure 3. Assume no unmeasured confounders

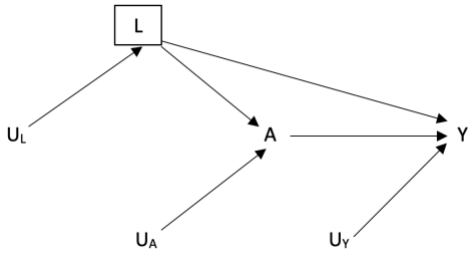

Figure 4. Assume independent unmeasured confounders

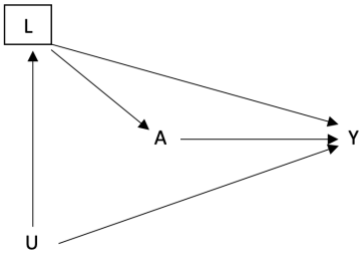

Figure 5. Assume no unmeasured confounders to A.

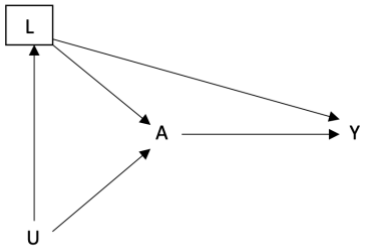

Figure 6. Assume no unmeasured confounders to Y

In the current study, any of the above assumptions on U for causal identification is strong. Unmeasured social, cultural, and contextual factors are likely to have an impact on A (e.g., youth's prosocial identity content), Y (reoffending behavior), and a subset of covariates L (e.g., socioeconomic status, risk level). This is common in observational studies though, as Robin and colleagues (2000) stated, "unbiased estimation by any method is impossible in the presence of unmeasured confounding factors without strong additional assumptions" (p.553). With the hope that residual confounding due to unmeasured confounders U may be small given that measured confounding from L on A can be adjusted by robust statistical methods, we may proceed with estimation on the effect of A and interpret the results with caution.

In the following sections, I describe using multiple logistic regression, TMLE, and MSM with an effect modifier and TMLE estimator to address the four research questions of the study. For each question, when causal modeling is applied, I operationalize the research question into a causal question, specify the MSM and its corresponding statistical model, define the casual parameters of interest and their statistical estimands, and discuss estimation approaches.

Analytical Approach for Research Question One

Question one is to examine whether other-oriented identity adds incremental utility to conventional identity, in protecting justice-involved youth from reoffending. Logistic regressions are conducted to address this question.

Assuming a linear log relation between outcome Y (reoffending) and predictors A (other-oriented identity and/or conventional identity), and assuming independence of observations in

the data, two nested multiple logistic regression models with other-oriented identity and conventional identity being added stepwise are used to test if adding any of the two variables improves model fit over another. For example, model (1) is written as:

$$log \text{ (odds of Y)} = \beta_0 + \beta_1 A_1 \quad (1)$$

where A_1 is conventional identity, which is a continuous variable measured by youth's sum of scores on four parcels (Honest/True, Integrity, Virtuous, Knows/Choose Right) in the Moral Ideal Self Scale; and Y is the outcome.

Model (2) is written as:

$$log \text{ (odds of Y)} = \beta_0 + \beta_1 A_1 + \beta_2 A_2 \quad (2)$$

where A_2 is other-oriented identity, which is a continuous variable measured by youth's sum of scores on the Loving/Caring parcel in the Moral Ideal Self Scale.

A likelihood ratio test for goodness of fit is conducted to examine if model (2) improves data fitting over model (1). Akaike Information Criterion (AIC) is calculated as well as another indicator of model fit. If significant improvement is not observed, then there is evidence that adding variable A_2 in the model does not substantially improve model fit. In other words, in such case other-oriented identity has no significant impact on predicting the outcome over conventional identity, thus it has no incremental utility in predicting youth's reoffending.

In these logistic regression models, I did not adjust for any covariates because the aims are 1) to identify which variable—either other-oriented or conventional identity—has a direct and significant relation with the recidivism outcome, and 2) to understand whether adding one to another improves the prediction of the outcome, without the influence of other variables.

Analytical Approaches for Research Questions Two, Three, and Four

The analytical approaches for research questions two, three, and four are similar. Here, I use research question two as an example to illustrate the approaches.

Question two is to examine the extent to which prosocial identity prominence (marker is moral internalization) moderates the relation between prosocial identity and youth's reoffending. Three different analytical approaches—TMLE analysis for the main effects of prosocial identity (measured by youth's moral ideal self) and prosocial identity prominence (moral internalization), respectively; multiple logistic regression with an interaction term; and MSM with an effect modifier and TMLE estimator—were conducted separately to address this question.

TMLE Analysis for the Main Effect of Holistic Prosocial Identity on the Recidivism Outcome.

As a relatively new approach that merges the strengths of the Inverse Probability of Treatment Weighted (IPTW) method (Robins et al., 2000) with machine learning techniques, TMLE is a double-robust, semi-parametric estimation procedure that approximates experimental conditions by creating comparison groups that are similar in their observed characteristics (van der Laan & Rose (2011). This approach takes account for the confounding effects from observed covariates (i.e., age, gender, race, SDI, and baseline risk) to estimate the effect of exposure A (e.g., prosocial identity measured by moral ideal self, coded as low or high level) on the outcome Y (i.e., recidivism, coded as a binary variable).

Estimating with a basic TMLE involves several steps. The first step is estimating the expected probability of outcome Y based on the exposure A and covariates L. This is referred to as the initial estimate of Y. The second step involves applying targeting steps, including the IPTW method, to update this initial estimate of Y. Then, both the updated and initial estimates of Y are used to calculate the parameters of interest, such as the difference in probabilities of occurring outcomes Y under different conditions of exposure A. During this process, machine learning techniques with cross-validation (e.g., SuperLearner) can be applied to obtain the initial estimate of Y and/or to the IPTW method. The IPTW method estimates each individual's probability of being in their observed condition *a* given their covariates *l* (i.e., the estimate of the treatment assignment mechanism). Incorporating machine learning techniques in TMLE offers several advantages, including avoiding assumptions about the functional form of statistical models and improving the accuracy of estimates. TMLE is a double-robust approach because if either the initial estimate of Y or the estimate of the treatment assignment mechanism is unbiased, the updated estimate of Y will be consistent (i.e., it converges to the truth as the sample size increases) and efficient (i.e., it has the smallest possible variance). If both the initial estimate of Y and the estimate of the treatment assignment mechanism (i.e., via IPTW) are unbiased, then the TMLE estimator will be most efficient.

Multiple Logistic Regression with Interaction

Assuming a linear log relation between outcome Y and predictors A, assuming independence of observations, and given that the correlations between predictors are weak to moderate (refer to the Results section: Psychometric Properties of Baseline Measures), a logistic regression model with an interaction term is expressed as follows:

$$log\ (\text{odds of } Y|L) = \beta_0 + \beta_1 A + \beta_2 L + \beta_3 V + \beta_4 AV \quad (3)$$

where A is prosocial identity, which is a continuous variable measured by youth's sum scores of all the 20 items in the Moral Ideal Self Scale. L is a set of covariates. And V is prosocial identity prominence measured by moral internalization and is a continuous variable. Coefficient β_1 estimates the conditional association between a one-unit change in A and the log odds of Y when V is zero and holding L constant. Coefficient β_3 estimates the conditional association between a one-unit change in V and the log odds of Y when A is zero and holding L constant. Coefficient β_4 estimates how much the conditional association between a one-unit change in A and the log odds of Y changes as V changes by one unit, holding L constant. Estimated coefficients and their *p* values provide evidence on whether the relation between a youth's moral ideal self and the youth's reoffending is influenced by the degree of the youth's moral internalization.

MSM with an Effect Modifier and TMLE Estimator

Another analytical approach to address question two is to apply a causal framework, under which the research question can be operationalized in causal language as: *how does the expected probability of reoffending vary as a function of youth's level of prosocial identity content and level of prosocial identity internalization?* With relations between U, L, A, and Y variables being specified in Figure 7, now the focus is on the interaction effect of A and V on Y, where V is prosocial identity internalization. Here both A and V were re-coded as binary indicating two levels (low or high) of the holistic prosocial identity content and internalization based on 50% percentile

of youth's scores on all items of the Moral Ideal Self Scale and Moral Internalization Scale, respectively.

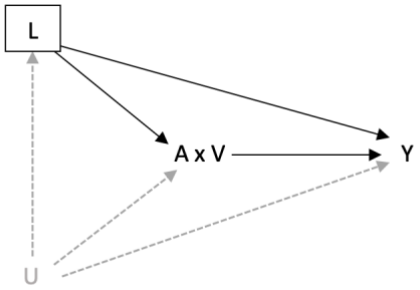

Figure 7. DAG for the interaction effect of A and V on Y, adjusting for L

To answer the research question, a working logistic marginal structural model with effect modifier V (Neugebauer & van der Laan, 2007; Robins et al., 2000) can be used to project the true causal curve between A, V and Y. The word "marginal" means to model on the marginal distribution of the counterfactual outcome Y_a (Robins et al., 2000). The word "working" means the model provides a summary measure of the true causal curve, which can be voluntarily mis-specified when one is only interested in the overall causal trend (Neugebauer & van der Laan, 2007). Effect modifier V is measured at baseline. Because V is of substantial research interest and the relation of A and Y is believed to differ by levels of V, it can be included in a working logistic MSM as:

$$E(Y_a \mid V) = \textit{logit}\ pr[Y_a = 1 \mid V = v] = \beta_0 + \beta_1 a + \beta_2 v + \beta_3 av \qquad (4)$$

where *a* is levels in A, *v* is levels in V, β_3 is the coefficient for interaction, and $\beta_1 + \beta_3 v$ represents the effect of condition *a* (e.g., a high level of prosocial identity content) on a linear logistic scale within level *v* (e.g., a high level of prosocial identity internalization) of covariate V. The parameters of interest are β_0, β_1, β_2, and β_3, and e^{β} is the causal odds ratio associated with different terms in the working MSM. Specifically, β_1 is the conditional effect of A on Y when V = 0, β_2 is the conditional effect of V on Y when A = 0, and β_3 estimates how much the difference in Y between two individuals that differ by one level on A changes as V changes by one level.

Under the assumption of no or minimal unmeasured confounding and that we have sufficient data support, if we adjust for measured confounders L (see Figure 9), unbiased parameters of model (4) can be estimated by fitting a statistical model in the observed data:

$$E'(Y_a \mid V) = \textit{logit}\ pr[Y = 1 \mid A = a, V = v] = \beta_0' + \beta_1' a + \beta_2' v + \beta_3' av \qquad (5)$$

where β' is the statistical quantity used to approximate β.

The TMLE estimator is used to adjust for confounders L and obtain estimates of β'. As introduced earlier, the MSM with the TMLE estimator has some key differences from traditional logistic regression. The MSM model focuses on counterfactual outcomes rather than observed outcomes. It adjuts for confounders L by incorporating them into the TMLE process, as opposed to adding L directly to the regression model. This approach is particularly beneficial in longitudinal settings where adding L to the regression model may introduce bias (Robins et al., 2000).

Values of coefficient β_1', β_2', β_3' in model (5) and their *p* values provide evidence on whether the relation between prosocial identity and youth's expected probability of recidivism would differ by levels of prosocial identity internalization.

Questions three and four employ similar analytical approaches to question two, where TMLE analysis for the main effects of the predictors, multiple logistic regression with an interaction term, and MSMs with TMLE are used. These methods examine the main effects of predictors of interest in a more rigorous way and test the extent to which prosocial identity validation (approximated by general self-efficacy), social support from peer and/or caregiver, and school connectedness moderate the relation between prosocial identity and youth reoffending. When applying a causal framework, the questions are operationalized as follows: *How does the expected probability of reoffending vary as a function of the youth's level of prosocial identity, and levels of self-efficacy, social support, or school connectedness?* For social support, I examined both peer and caregiver support separately and combined.

Appendix D: Results from Preliminary and Primary Analyses

Table 1. Correlations between independent variables and age, SDI, and baseline risk (*N*=760)

	Age at Intake	SDI score	Risk Score
Moral Ideal Self	0.04	-0.05*	-0.18***
Other-oriented Identity	0.00	-0.07**	-0.18***
Conventional Identity	0.06*	-0.04	-0.16***
Moral Internalization	-0.01	-0.10***	-0.18***
Self-efficacy	0.08***	-0.01	-0.10***
Social Support (Peer & Caregiver)	-0.09***	-0.02	-0.20***
Peer Support	-0.06*	-0.04	-0.10***
Caregiver Support	-0.09***	-0.002	-0.19***
School Connectedness	-0.07**	-0.11	-0.20***

** $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$*
Note: test statistics is Kendall's τ

Table 2. Differences on measures of independent variables between gender groups (*N* = 760)

	Male (*N*= 533) Mean (SD)	Female (*N*= 227) Mean (SD)	Point-biserial Correlation r_{pb}	Kruskal-Wallis Effect Size η^2
Moral Ideal Self	5.83 (0.86)	6.01 (0.79)	0.10**	0.008
Other-oriented Identity	5.89 (0.89)	6.11 (0.81)	0.11**	0.01
Conventional Identity	5.78 (0.93)	5.92 (0.89)	0.07*	0.004
Moral Internalization	3.98 (0.69)	3.97 (0.73)	-0.01	
Self-efficacy	30.00 (4.45)	29.56 (4.60)	-0.04	
Social Support (Peer & Caregiver)	4.11 (0.71)	4.04 (0.68)	-0.05	
Peer Support	3.94 (1.08)	3.90 (1.14)	-0.02	
Caregiver Support	4.16 (0.77)	4.08 (0.74)	0.05	
School Connectedness	3.45 (0.89)	3.21 (0.88)	-0.12***	0.008

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 3. Differences on measures of independent variables between racial groups (*N*=760)

	White (*N*=276) Mean (SD)	Black (*N*=350) Mean (SD)	Latinx (*N*=102) Mean (SD)	Other (*N= 32)* Mean (SD)	Kruskal-Wallis Test		Pairwise Comparison
					p	η^2	*(*adjusted *p)*
Moral Ideal Self	5.97 (0.79)	5.85 (0.86)	5.70 (0.93)	6.08 (0.66)			
Other-oriented Identity	6.07 (0.80)	5.90 (0.89)	5.76 (1.03)	6.25 (0.21)	*	0.01	W vs. L* W vs. B* O vs. L*
Conventional Identity	5.89 (0.88)	5.81 (0.96)	5.66 (0.93)	5.94 (0.87)			
Moral Internal-ization	4.13 (0.66)	3.89 (0.71)	3.82 (0.70)	4.19 (0.65)	***	0.03	W vs. L*** W vs. B*** O vs. L* O vs. B*
Self-efficacy	29.63 (4.26)	30.08 (4.75)	29.54 (4.57)	30.66 (3.03)			
Social Support (Peer & Caregiver)	4.13 (0.65)	4.07 (0.73)	4.06 (0.70)	4.00 (0.86)			
Peer Support	4.09 (0.99)	3.80 (1.16)	3.98 (1.05)	3.97 (1.23)	*	0.01	W vs. B*
Caregiver Support	4.15 (0.70)	4.15 (0.78)	4.08 (0.76)	4.10 (0.95)			
School Connect-edness	3.53 (0.87)	3.27 (0.90)	3.35 (0.86)	3.31 (0.97)	**	0.01	W vs. B*

** $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$*

Table 4. Multiple logistic regressions with social support as the main predictor and adjusting for different sets of covariates (*N*=760)

Variable	Any Petition within One Year					
	OR	95% CI	OR	95% CI	OR	95% CI
Social Support (Peer & Caregiver)	0.96*	0.93, 0.99	0.96*	0.93, 0.99	0.98	0.94, 1.01
Age at Intake	1.17	1.00, 1.37	1.17	0.99, 1.38	1.13	0.96, 1.34
Female	0.60*	0.35, 0.98	0.61	0.36, 1.00	0.62	0.36, 1.02
Race-Latinx	1.38	0.59, 3.01	1.04	0.43, 2.40	0.97	0.39, 2.26
Race-Black	3.31***	1.99, 5.74	2.59**	1.48, 4.71	2.14*	1.20, 3.93
Race-Other/unsure	0.82	0.13, 3.05	0.75	0.12, 2.82	0.62	0.09, 2.38
SDI Score			1.01*	1.00, 1.02	1.01	0.99, 1.01
Baseline Risk Score					1.07***	1.04, 1.11
Intercept	0.04*	0.00, 0.62	0.01**	0.00, 0.28	0.01**	0.00, 0.28

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 5. MSM analyses with TMLE estimator integrated with Super Learner, by prosocial identity (PI) content and social support, using age, gender, race, SDI, and baseline risk as covariates

	Any Petition within One Year (*N*=760)	
Variable	OR	95% CI
High PI Content (High Moral Ideal Self)	0.65	0.34, 1.25
High Social Support (Peer & Caregiver)	0.41**	0.21, 0.79
High PI Content x High Social Support	1.93	0.74, 5.06
Intercept	0.23***	0.17, 0.31
High PI Content (High Moral Ideal Self)	0.79	0.42, 1.51
High Caregiver Support	0.54*	0.29, 1.00
High PI Content x High Caregiver Support	1.24	0.49, 3.18
Intercept	0.21***	0.16, 0.29

** $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$*