Check for updates

Original Article

International Criminal Justice Review
1-28
© 2022 Georgia State University

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10575677221082071
journals.sagepub.com/home/icj

Does Contact with the Justice System Influence Situational Action Theory's Causes of Crime? A Study of English and German Juveniles

Florian Kaiser[1]

Abstract

To explore why system contact often has no crime-preventative effect, the current study examined the effects of juvenile justice contact on Situational Action Theory's (SAT) causes of crime, including personal morals, deviant peer associations, and detection risk perceptions. The analysis is based on a sample of English (Peterborough Adolescent and Young Adult Development Study) and German (Crime in the modern City study) juveniles. Propensity score matching was applied to estimate whether the lenient system contacts influenced the causes of crime in the year after the contact. The treatment effect estimates are mostly insignificant and relatively small. The few significant estimates in the English sample suggest that official contact slightly increased deviant peer associations and decreased feelings of moral guilt. Overall, the findings suggest that system contact may often have no crime-preventative effect as it does not (Germany), or only slightly (England) affect SAT's causes of crime. Previous studies, primarily based on the U.S. data, often reported more substantial effects that mostly operated in a crime-amplifying direction. It is speculated whether the less substantial impact in the current study can be attributed to the overall more lenient, diversion-oriented handling of the examined English and German offenders.

Keywords

effects of juvenile justice system contact, causes of crime, Situational Action Theory, propensity score matching, cross-national research

Introduction

A primary objective of modern criminal justice systems is to prevent crime. If the systems fail to achieve this goal and a person breaks the law, legal actors (e.g., police officers) typically seek to prevent further offending by arresting and, if necessary, sanctioning that person. In recent

[1] Institute of Criminal Law and Criminology, University of Münster, Münster, Germany

Corresponding Author:
Florian Kaiser, Institute of Criminal Law and Criminology, University of Münster, Bispinghof 24/25, 48143 Münster, Germany.
Email: florian.kaiser@uni-muenster.de

decades, numerous studies have examined whether such contact with the criminal justice system (including, e.g., apprehension, arrest, diversion, and sanctions) actually prevents reoffending. Contrary to the goals of criminal justice agents, the majority of these studies suggest that system contact is either relatively ineffective or even increases criminal involvement (for reviews, see Barrick, 2014; Bernburg, 2019; Huizinga & Henry, 2008; Kleck & Sever, 2017).

Using data that resulted in insignificant findings in a previous investigation (Boers et al., 2022), the current study addresses the following question: Why do contacts with the criminal justice system often have no crime-preventative impact? Answering this question theoretically and empirically requires focusing on the intervening factors that may mediate the effects of official contact on reoffending.

Criminologists have identified numerous factors as theoretically relevant mediators. *Perceptual deterrence* theorists assume that official contact may reduce reoffending by increasing sanction threat perceptions (Paternoster, 2018). *Labeling* scholars suggest that official contact may amplify reoffending (1) by initiating or facilitating the development of a deviant self-concept, (2) by initiating or increasing the association with deviant peers, or (3) by inhibiting the punished offenders' social bonds and life chances (Bernburg, 2019). The *procedural justice* theory stresses that the effect of system contact depends on whether apprehended offenders feel that the proceeding against them was fair and just and whether they consequently view the law and its enforcement as legitimate (e.g., Slocum et al., 2016). *Defiance* theory emphasizes that the impact of punishment depends on how offenders perceive the sanction (e.g., unfair and stigmatizing), how strongly they are bonded to the sanctioning agent and community, and whether they subsequently develop feelings of shame or self-righteous anger (Sherman, 1993).

Confronted with such an extensive (but not exhaustive) list of potentially relevant intervening factors, one wonders which are the most relevant mediators. So far, most empirical studies did not concentrate on this question. They instead tested the presumptions of a single theory, such as deterrence theory (e.g., Anwar & Loughran, 2011; Pogarsky et al., 2005) or labeling theory (e.g., Bernburg et al., 2006; Wiley et al., 2013). However, the empirical literature indicates that mechanisms proposed in different theories may be at work simultaneously. Hence, it may be better to integrate these mechanisms into a more general theoretical framework. The need for such integration was recently highlighted by Piquero et al. (2011, p. 338), who stressed that "it may be more profitable to think of a general theory of sanctions rather than deterrence, labeling, or defiance theory." When considering the requirements for such a general theory of sanctions, it is apparent that it must clearly define the factors that directly cause crime. It is only when a sanction reduces these key causal inputs that it can ultimately reduce reoffending. By distinguishing between direct causes of crime and more distal factors, a general theory should be able to "separate the wheat from the chaff."

All theories that clearly define the causes of crime are thus candidates for a starting framework for a general theory of sanctions. The current study selects Wikström's *Situational Action Theory* (SAT) as a promising starting point (Wikström et al., 2012). SAT is chosen because it has performed well empirically so far (Pauwels et al., 2018)[1], provides a sophisticated action model, and goes to great lengths to distinguish between direct antecedents of criminal offending and more distal factors. In particular, the theory differentiates the *causes of crime* from the *causes of the causes* (Wikström, 2011). *Causes of crime* are the few factors that directly influence criminal involvement. *Causes of the causes* are, in contrast, factors that have only an indirect effect on criminal offending through their influence on the causes of crime. By providing this terminology, SAT allows identifying the causes of crime as the crucial mediators of system contact effects. Some of these causes of crime coincide with intermediate factors outlined in contemporary versions of deterrence and labeling theory. Using SAT as starting framework, the current study can thus integrate some ideas of the latter theories about sanctioning effects. In doing so, it is—at least to my knowledge—the first empirical application of SAT to study the impact of criminal justice interventions.[2]

The current study aims to shed some more light on why a relatively large number of previous studies—among them one that used the data at hand—found no crime-preventative effects of system contact. It does so by empirically examining to what extent a juvenile justice contact influences the antecedents of crime as postulated in SAT. The analyses are based on data from two criminological panel studies conducted in England and Germany. So far, nearly all studies that examined official contact effects on antecedents of crime were conducted with the U.S. samples. While experts already consider this study base to be too small (e.g., Bernburg, 2019; Farrington & Murray, 2014; Huizinga & Henry, 2008; Paternoster & Iovanni, 1989), research outside the United States can be described as extremely sparse (for exceptions, see Murray et al., 2014; Schulte, 2019; Zhang & Messner, 1994)[3]. This lack of research is unfortunate because system contact in different juvenile justice systems (with their various policies, sanctions, and police behavior) may produce different effects (see Huizinga et al., 2003). The German and English systems offer an insightful research context. They are much more lenient than the U.S. system but still characterized by some differences like in police behavior. Hence, it is worth exploring whether results obtained with the U.S. samples can be generalized to the English and German contexts.

Intervening Factors in the Framework of SAT

This section shows how SAT can provide a starting framework for a general theory of sanctions. For this purpose, it first introduces how contemporary versions of labeling and deterrence theory are applied to study sanctioning effects. By highlighting some theoretical weaknesses of these applications, the section then shows how SAT attempts to overcome these problems and may integrate some ideas of the classic theories.

Rooted in symbolic interactionism, labeling theory assumes that humans are creatures that continuously construct meaning of the environment and themselves by interpreting significant symbols in social interactions. As a result of the meaning people attach to official contact, it may trigger three criminogenic processes according to contemporary versions of labeling theory (Bernburg, 2019; Krohn & Lopes, 2015): (1) the development of a deviant self-concept, (2) the reduction of conventional social bonds and life chances, and (3) the involvement in deviant peer groups. These processes are also studied and typically supported in recent empirical applications of labeling theory (e.g., Bernburg et al., 2006; Krohn et al., 2014; Wiley et al., 2013). However, from my point of view, this labeling research is theoretically underspecified. It typically only defines a set of "orienting" criminogenic processes triggered by societal reaction to deviance. It lacks a uniting explanation in the form of an action model which specifies how the changes (e.g., in self-concept) brought about by these processes ultimately operate (together) to explain an increase in deviant or criminal behavior.

Modern deterrence theory, in contrast, borrows its action model specification from rational choice theories. Consequently, deterrence research assumes that humans are self-interested and rational beings that actively choose one action alternative over others as a result of cost-benefit considerations (Paternoster, 2010, 2018). Recent sanctioning research typically tests perceptual versions of deterrence theory (e.g., Lochner, 2007; Matsueda et al., 2006; Pogarsky et al., 2005). These versions assume that an intervention can affect future criminal behavior only through its impact on the perceptions of punishment certainty, severity, or celerity. Due to its recourse to rational choice frameworks, deterrence theory allows—other than labeling theory—for precise implications about how formal control ultimately leads to changes in delinquency via changes in sanction threat perceptions. However, standard rational choice models have been immensely criticized over the last decades. Critics primarily noted that these approaches view humans as overly rational and calculative beings. Consequently, they cannot explain automatic spontaneous, habitual, or emotionally driven behavior, which may make up a large part of human behavior (e.g., Kroneberg et al., 2010; Wikström & Treiber, 2016b).[4]

Partly as a response to these problems of existing criminological theories, Per-Olof Wikström introduced SAT at the beginning of the 2000s (Wikström et al., 2012; Wikström & Treiber, 2016b). One of his primary goals was to specify a proper action model which explains why an act of crime is (not) committed. To do this, SAT's action model attempts to consider modern insights from the behavioral sciences, among them the finding that human (including deviant) behavior is often habitual. Taking the perspective that humans are rule-guided (instead of primarily self-interested) beings, SAT's action model has fared well in empirical tests so far (Pauwels et al., 2018). As it also acknowledges the importance of some of the concepts included in deterrence theory (e.g., perceptions of consequences) and labeling theory (e.g., deviant peers), it seems a promising framework to integrate some of the mechanisms proposed for decades in these classic contenders.

Before discussing this integration in more detail, however, I will first introduce the basic ideas of SAT and thereby extract the causes of crime as the article's outcome variables of interest (for a more detailed introduction to SAT, see Wikström et al., 2012). In its action model, SAT describes how particular characteristics of a person interact with inducements of the setting (i.e., the immediate environment) to create motivations (temptations and provocations) and initiate and guide a perception–choice process that finally produces criminal behavior.

The main causal factors in the perception process are an *individual's personal morals* and the *setting's moral norms*. While personal morals are a person's "value-based and emotionally grounded views about what is the right or wrong thing to do or not to do in particular circumstances," a setting's moral norms are the "perceived shared rules of conduct and their degree of homogeneity" (Wikström, 2020, p. 193). The combination of the two factors determines the action alternatives an individual perceives (Wikström & Treiber, 2016b). Ultimately, SAT predicts that individuals are more likely to perceive (and choose) crime as an action alternative in circumstances in which their own morals and the perceived moral norms of the setting are more affirmative of crime.

After perceiving particular actions as potent alternatives, the choice of action can typically proceed in two different modes (Wikström & Treiber, 2016b): choices can be made in a habitual or in a more deliberative manner. Habitual decisions are typically made in familiar circumstances in which the personal morals and setting's moral norms are congruent (e.g., both are crime-affirmative) because individuals then normally perceive only one viable course of action. Choices are, in contrast, typically deliberative if a person acts in unfamiliar circumstances, if there is a conflict between a person's morals and the setting's moral norms, and if opportunities for reflection are not undermined by, for example, time pressure or strong emotions. In this circumstance, individuals have time to deliberate over multiple action alternatives. According to SAT, it is only in this more deliberative mode of decision-making that additional factors—beyond personal morals and the setting's moral norms—have a causal influence on criminal behavior.

SAT suggests that these additional causal factors, namely the *ability to exercise self-control* and the *perceived risk of consequences*, exert their causal impact within two so-called control processes (Wikström & Treiber, 2016b). As an internal control process, self-control refers to acting according to one's own morals despite external incentives or pressure to do otherwise. The *ability to exercise self-control* depends on momentary (e.g., intoxicated) and dispositional (e.g., executive functions) factors. SAT generally assumes that the higher the self-control ability of people, the more likely they will resist external criminogenic influences (e.g., provided by deviant peers) and act law-abiding. SAT's external control process is called *deterrence*. Deterrence describes "the avoidance of breaking a moral rule (committing an act of crime) because of the fear[5] of consequences" (Wikström, 2008, p. 347). The likelihood that individuals are deterred increases with their perceptions of how likely and severe consequences are when committing a crime. These perceived consequences depend on deterrent cues provided by the setting (e.g., by authority figures) and on how sensitive a person is to these cues, that is, how likely the person processes them so that they perceive

them as punishment threats (Wikström, 2008). Generally, SAT assumes that the higher the *perceived threats of punishment*, the more likely a person will abide by the law despite their urge to commit a crime (Wikström et al., 2011).

Overall, SAT's action model thus suggests that the direct causes of crime operate in a person-environment interaction to produce moral (including criminal) action (Figure 1). The causes of crime that form a person's *crime propensity* are their personal morals and self-control capabilities. Deterrence sensitivities (or sanction threat perceptions) can be regarded as an additional personal cause of crime (see Wikström et al., 2011). SAT furthermore suggests that particular places (e.g., the city center) and the exposure to deviant peers are major environmental causes of crime (e.g., Wikström et al., 2012). These two factors contribute to the criminogeneity of the immediate environments a person takes part in and thus to their *criminogenic exposure* (i.e., the criminogeneity of the setting's moral norms and enforcement). Especially deviant peers have been found to be a critical facilitator of criminogenic exposure by providing (and enforcing) crime-affirmative moral norms (see Beier, 2018; Kaiser, 2021; Wikström et al., 2012).

How can a criminal justice contact now influence delinquency? According to SAT, formal control cannot directly impact criminal involvement but can change it only indirectly via changing the causes of crime. As a cause of the causes, official contact must trigger the following mechanisms specified in SAT's recently refined Developmental Ecological Action model (DEA model; Wikström, 2020): (1) psychosocial processes and (2) socioecological processes. The two key psychosocial processes that influence a person's crime propensity are moral education and cognitive nurturing. While moral education is a learning and evaluation process responsible for changes in an individual's personal morals, cognitive nurturing describes an experiential process that influences the ability to exercise self-control. The socioecological processes of self-selection and social selection, on the other hand, are responsible for changes in a person's criminogenic exposure. While self-selection refers to a person's choices to participate in particular settings (including particular people), social selection refers to how cultural and structural conditions in a jurisdiction enable or restrict the access of particular people (and thus also their self-selection) to particular settings.

However, since SAT has so far not been applied to study the impact of criminal justice interventions, the developmental processes are somewhat general or abstract in this regard. The DEA model,

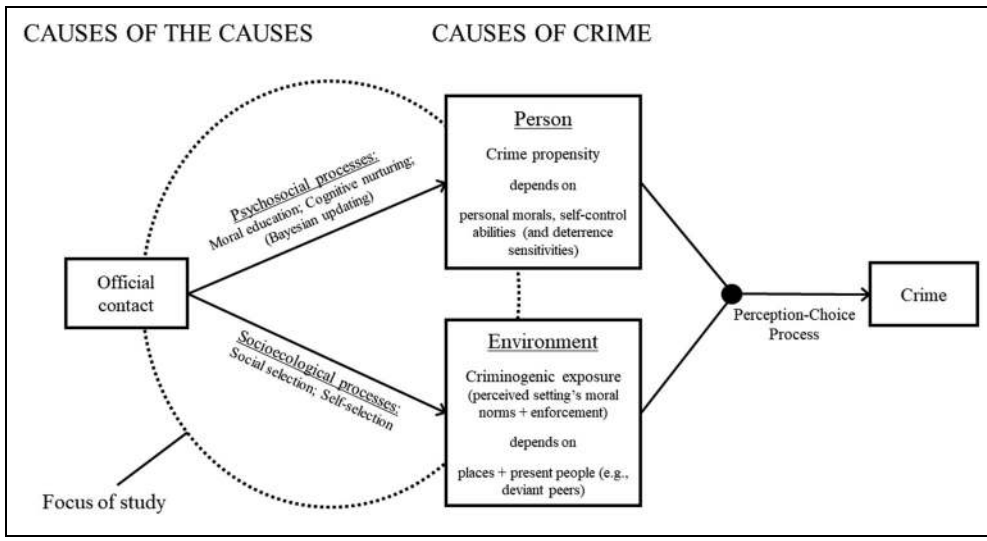

Figure 1. Situational Action Theory (SAT) and the impact of official contact (see also Wikström, 2020).

furthermore, does not explicitly specify a mechanism affecting deterrence sensitivities (or sanction threat perceptions), probably because these typically play only a minor role in SAT introductions. Integrating ideas of classic sanctioning theories may help to overcome some of these problems, as these have specified detailed mechanisms triggered by formal control over decades (including processes influencing sanction threat perceptions). They, thus, seem a valuable source to enrich the processes defined in the DEA model. Accordingly, the following subsections will consider ideas of the DEA model and the classic theories when discussing how official contact may affect some of SAT's key causes of crime. In particular, the presented processes will be related to personal morals, deviant peer associations, and detection risk perceptions since reliable, comparable data are available only for these factors (which are among the mainly tested causes of crime in previous SAT research; Pauwels et al., 2018).

Official Contact and Personal Morals

Since the classic sanctioning theories have not specified any processes relevant to personal morals, the current study will rely on the DEA model to infer implications regarding how official contact may change this cause of crime. The DEA model suggests that people typically acquire their morals gradually and change them through a continuous process of moral education (see Wikström, 2020; Wikström & Treiber, 2016a, 2018). Moral education, as one example of social learning, has three sub-mechanisms: (1) instruction, (2) observation, and (3) trial and error. *Instruction* takes place if moral instructors (e.g., parents or police officers) actively "communicate information about the rules of conduct which apply to different contexts" (Wikström & Treiber, 2016a, p. 80). *Observation* means that a person keeps track of others' actions and the action-induced consequences (e.g., rewards and punishments). The *trial and error* process refers to a person's own experimentation with actions and their consequences. Deterrence experiences, that is, "experiences of threats of punishment, and punishments" (Wikström, 2008, p. 356), are one form of experience that influences an individual's morals. As one particular type of deterrence experience, official interventions can be described as *instructive* reactions of legal actors to the "*trial-and-error*-experimentation" of an offender.

SAT does not assume that official contact with the criminal justice system increases or decreases personal morals on average. It instead implies that the effect depends on the specific ways others react to a criminal offense and subsequent official contact, and on how the offender evaluates her or his crime and sanctioning experience as well as the formal and informal reactions of others. In particular, the moral education process depends on such factors as (1) the consistency of reactions and instructions, (2) the offender's internal evaluation of these instructions or reactions, (3) the offender's prior experiences and existing personal morals, and (4) the offender's psychosocial machinery (e.g., cognitive capabilities; Wikström & Treiber, 2016a, 2018).

Formal reactions may influence a person's morals only substantially when they initiate changes in their day-to-day moral experiences in interaction with significant others (e.g., parents or peers). That is because daily activities and interactions with significant informal others play a continuous role in an individual's moral development, while rare encounters with the police or prosecutors typically do not (Wikström, 2008). Before youths have official contact with the criminal justice system, they are likely to have already had several crime-related experiences, including (moral and deterrence) experiences with the reactions of, for example, peers or parents. As a result, these youths' morals are likely to be relatively consolidated and, as a consequence, less malleable by new experiences. Their moral learning contexts have probably already been formed and solidified by previous reactions to criminal activity and are thus unlikely to change substantially due to new reactions. If the moral learning context is relatively fixed, the moral content learned within it should remain relatively stable because a person's morals only develop and change in interaction with the environment

(Wikström, 2008). However, if the social context changes due to a system contact (e.g., a person gets more involved with deviant peers), personal morals may also change (e.g., may get more crime-affirmative) through the altered continuous moral experiences within these modified contexts.

The impact of official contact on personal morals has been largely neglected as an object of empirical investigation (Wikström, 2008). Only recently have Wiley and colleagues devoted part of their work to this topic using data from the *Gang Resistance Education and Training program* (Slocum et al., 2016; Wiley, 2015; Wiley et al., 2013, 2017; Wiley & Esbensen, 2016). Matching juveniles who had police contact to juveniles who had no contact but were otherwise similar (e.g., in their pre-contact morals), they found that those with contact subsequently reported lower personal morals (measured by the increased adherence to street codes and less anticipated guilt for deviant behavior) than their matched counterparts. The effects were generally stronger among individuals who had been arrested than among those who had only been stopped by the police (Wiley et al., 2013; Wiley & Esbensen, 2016). Besides Wiley and colleagues' investigations, other—less methodologically sophisticated—analyses found similar results (Ageton & Elliott, 1974; Kaplan & Johnson, 1991; Schulte, 2019).

Official Contact and Deviant Peer Associations (Criminogenic Exposure)

To discuss potential changes in deviant peer associations, I rely on ideas established over several decades by labeling theorists, who have identified mechanisms that may lead from official contact to an individual's exclusion from conventional networks and inclusion in deviant peer groups. These mechanisms can be seen as potential manifestations of the more general self-selection (and social selection) processes specified in the DEA model (Figure 1).

Labeling theorists outline the following processes that lead to the association with deviant peers (Bernburg, 2019; Paternoster & Iovanni, 1989): In the first step, prosocial others (e.g., family, conventional peers, and employers) may—due to their knowledge of the crime or system contact—vilify the offender and turn away from her or him. Second, the offender may actively withdraw from conventional others to avoid anticipated adverse reactions to the crime and the system contact. Both processes lead to exclusion from conventional society and socially isolate the individual. In the third step, the individual who came into contact with the criminal justice system may become involved in a deviant peer group, often because of her or his actual or perceived isolation. Labeling theory assumes that the apprehended offender may befriend individuals who share the same offending and sanctioning experiences because she or he hopes for acceptance or even admiration among these new friends.

Labeling theory, however, also acknowledges that official contact may have no impact or may even weaken deviant peer associations (Bernburg, 2019; Paternoster & Iovanni, 1989). Negative reactions of parents or conventional friends to a crime may, for example, warn the offender that future offending could alienate them (further). When the offender is ashamed in front of significant others (e.g., parents or partners) or fears losing them due to further delinquency, she or he may decide to stop spending time with deviant friends to prevent them from providing opportunities for and exerting pressure to engage in further rule-breaking. As a consequence of official contact, an offender may also be monitored more closely by law-abiding others, who may actively (try to) restrict the offender's exposure to criminogenic settings or persons (e.g., deviant peer groups). Furthermore, labeling theory assumes that, while the first system contact strongly influences sanctioned individuals, official contact effects diminish the more a person experienced labeling and stigmatization before. Thus, if previous informal and formal deterrence experiences have already led to an association with deviant peers, further system contacts are likely to have significantly less or no further influence on those associations.

Despite the importance of deviant peer groups as one of the strongest predictors of criminal involvement, only a few longitudinal studies have investigated the influence of official contact on

the subsequent association with deviant peers, peer groups, or gangs. Most of the results indicate an increased involvement with deviant friends or gangs among juveniles who previously had (more frequent or severe) contact with the juvenile justice system compared with juveniles who had less or no system contact (Bernburg et al., 2006; Johnson et al., 2004; Kaplan & Johnson, 1991; Krohn et al., 2014; Restivo & Lanier, 2012; Schulte, 2019; Wiley et al., 2013; Wiley & Esbensen, 2016). A related line of research, furthermore, found some evidence that official contact also had the potential to isolate offenders from conventional friends (e.g., Wiley et al., 2013; Zhang, 1994; Zhang & Messner, 1994)—which, according to labeling theory, is a precondition for the association with deviant peers.

Official Contact and Detection Risk Perceptions

Probably because SAT introductions typically grant deterrence sensitivities (or perceptions of punishment threats) no prominent spot, the DEA model does not present a process that may lead to changes in these sensitivities. Deterrence theorists, in contrast, have put great effort into explaining how individuals modify their risk perceptions (or risk assessments, which can be seen as an indicator of one's deterrence sensitivities; Wikström, 2008) after contact with the criminal justice system. The current study, thus, relies on their ideas when reviewing the so-called *Bayesian updating model* (Anwar & Loughran, 2011; Kreager & Matsueda, 2014). This model states that people will gradually adjust their prior (i.e., initial or existing) perceptions (according to probability rules) when exposed to new evidence or information related to the perception at hand. Individuals should, for example, decrease their risk perceptions when they or others commit a crime and get away with it, but should increase them when they or others are apprehended or sanctioned for committing a crime.

The Bayesian updating model, furthermore, assumes that the extent of these perception adjustments depends on the prior (criminal and sanctioning) experiences of the individual. The more prior information individuals possess relevant to their perception, the less weight the new experience or information has in modifying the perception. Suppose an action (crime) is novel for individuals. In that case, they must rely mainly on the actual experience to infer future consequences (e.g., apprehension and sanction) of similar actions since there is no or only little former experience on which to base their inferences. Consequently, this recent experience has a relatively large impact on the modification of the individuals' perceptions. Among individuals with more (criminal and sanctioning) experience, in contrast, the recent incident has less weight in adjusting a relevant perception because this perception is much more strongly informed by prior experiences.

There is growing longitudinal research on how an individual's own or significant others' apprehension or punishment experiences change her or his perceptions of detection or arrest risk (e.g., Anwar & Loughran, 2011; Lochner, 2007; Matsueda et al., 2006). Results indicate that people are generally "capable Bayesians" (Paternoster, 2010, p. 809). If they were (frequently) caught when committing a crime, they modified their detection risk perceptions on average upwards. If they were not (or infrequently) caught, they lowered their perceptions. The reported effects were generally small to modest, and some other studies found more mixed, inconclusive, or even conflicting evidence (e.g., Pogarsky et al., 2005; Schulte, 2019; Schulz, 2014).

Additional investigations suggest that the extent of updating depends on the criminal history of the individual (Anwar & Loughran, 2011). For offenders with a larger ratio of current crimes (i.e., crimes committed in the period of interest) to past crimes (i.e., crimes committed in the time before the period of interest), risk perception modifications after apprehension were larger than for those with a smaller current-to-past crimes ratio. Experienced offenders seem to need a stronger signal (i.e., a high arrest-to-crime ratio for many crimes) within a specific period to update their risk perceptions as strongly as novice offenders.

The English and German Juvenile Justice Systems

The theoretical ideas above indicate that more information is needed on the nature and certainty of system contacts to predict their impact on the juvenile population under investigation.

Generally, the legal framework of the English and German juvenile justice systems differed in several important ways in the early 2000s (i.e., in the current study period; Boers et al., 2022). The more punitive English system was characterized, for example, by an earlier age of criminal responsibility, fewer diversion possibilities, and longer durations of custodial sanctions. These differences in punitivity are also reflected in nation- and countywide statistics on the reactions of the juvenile justice systems. German juvenile offenders who had contact with the justice system were much more likely to be diverted than English offenders. English offenders, in contrast, were much more likely to be convicted to long-term custodial sentences than their German counterparts.

However, these differences in punitivity were much smaller in the samples that were analyzed in the present study. Most of the 15-year-old respondents experienced quite similar and relatively lenient treatment during the period of interest (see Table 1). Most cases were diverted out of the formal court system for informal handling, and the remaining juveniles were almost exclusively given educational measures (mainly unpaid community service).

Because the legal reactions were so similar in both samples, the main differences in the treatment experience of respondents with official contact were probably due to the divergent police practices in the two countries. English police acted more intrusively than their German counterparts when apprehending an offender, even when the individual was apprehended for a relatively minor offense (which made up most offenses in the current study). Officers usually arrested the juvenile and took her or him to the police station, where the offender's primary caregiver was called to pick her or him up. German police officers, in contrast, mainly recorded the offender's personal data in such a minor case and released the individual. In case of a standard diversion, the German prosecutor usually simply sent a formal letter informing the offender and principal caregivers of the decision, whereas in England, the decision (reprimand/warning) was delivered at the police station by a police officer in uniform after the offender's admission of guilt (Dünkel & Heinz, 2017; Home Office & Youth Justice Board, 2002).

English police officers acted not only more intrusively but were also encouraged to act more proactively than their German colleagues between 2002 and 2007. This proactive policing style was due to a short-lived policy change brought about through the *Offenses Brought to Justice Target* (OBJT). The OBJT required the justice system to increase the number of offenders who received a formal

Table 1. Type of Formal Reaction and Offenses the Juveniles Were Recorded for.

Formal reaction	ENG	GER	Offense type	ENG	GER
Diversion	73.0%	80.7%	Violent offenses	21.6%	24.1%
Conviction	27.0%	19.3%	*Offenses against the person*	*18.9%*	*20.7%*
Noncustodial measures	*24.3%*	*15.9%*	*Robbery*	–	*3.4%*
Short-term detention	–	*3.4%*	*Sexual offenses*	*2.7%*	–
Long-term imprisonment	*2.7%*	–	Property offenses	37.8%	43.7%
			Theft and handling	*24.3%*	*18.4%*
			Aggravated theft/burglary	*10.8%*	*11.5%*
			Fraud	*2.7%*	*13.8%*
			Vandalism	13.5%	1.1%
			Motoring offenses	2.7%	26.4%
			Drug offenses	5.4%	2.3%
			Other (weapons, threats)	18.9%	2.3%

Note: ENG = English sample (PADS+); GER = German sample (CrimoC); n(PADS+) = 37; n(CrimoC) = 88.

reaction by 20% within 5 years (Office for Criminal Justice Reform, 2004). To meet this target, the police acted more proactively than before and actively searched for crimes, which they found especially among juveniles who tended to be involved in more easily detectable delinquency (e.g., Bateman, 2017; Flanagan, 2007). This policy change led to the criminalization of many first-time, low-level juvenile offenders, whose cases previously mainly were dealt with informally (e.g., through a stern talking-to). Accordingly, the number of recorded first-time offenders rose by 53% in Cambridgeshire, and the arrest rate increased by 48.2% between 2002 and 2006 (Ministry of Justice, 2010). German officers, in contrast, generally responded only to those (youth) crimes that were reported to the police (Albrecht, 2010).

Hypotheses

Previous empirical research suggests that contact with the criminal justice system affects SAT's causes of crime. The findings indicate that system contact decreases personal morals and increases deviant peer associations and risk perceptions. These empirical results generally align with predictions that can be derived (under particular circumstances) by mechanisms postulated in SAT, labeling theory, or deterrence theory.

However, the theories mentioned above also acknowledge that the effects of official contact depend on individuals' previous crime-related experiences not only with formal control agents (e.g., the police) but also within their informal social network (e.g., with parents or peers). They imply that the more personal morals, peer associations, and risk perceptions were formed by previous (consistent) experiences, the less malleable they are.

Past experiences may be much different in the English and German samples than in previous studies that were based mainly on the U.S. samples. The U.S. juvenile justice system generally acts (through the police) in a more proactive way in searching for crimes, and the subsequent sanctions are more punitive than in the English and German systems (Howell et al., 2013; Huizinga et al., 2003; Snyder & Sickmund, 1999). Therefore, juveniles in the United States who have had contact with the criminal justice system may, on average, due to their higher risk of detection when committing crimes, not only have less experience with delinquency than their counterparts in England or Germany at the time of the official contact but may also be subjected to more intrusive interventions.

To derive a more informed hypothesis on the impact of criminal justice contact, one must thus consider the criminal and sanctioning history of the treated individuals (i.e., of those who had contact with the juvenile justice system) in the English and German samples (see online Supplemental Material S1). Information about this history indicates that 15-year-old delinquents who had official contact with the criminal justice system were indeed already typically quite experienced before this contact. That is, they reported a relatively high number of acts of (undetected) crime before their system contact.

In most cases, the individuals committed these crimes together with friends, and it is not unrealistic to assume that parents or teachers were also aware of at least some of their past offenses. Therefore, most youths probably had quite a few (informal) deterrence experiences before the investigated system contact. In this process, they already learned how likely or unlikely it is to be caught and how significant others in their social environment typically react to criminal activity. These experiences had already shaped and consolidated their morals, social associations, and risk perceptions, making them less malleable. In accordance with these learning processes, treated juveniles reported weaker personal morals and risk perceptions as well as stronger deviant peer associations than their untreated counterparts before their official contact (see section *Pretreatment covariate balance*).

The theories assume that an intensive intervention or a high detection or arrest certainty would be necessary to change such consolidated causes of crime. The severity and certainty of sanctions, however, were relatively weak among the individuals under investigation. On the one hand, they

were apprehended only once or twice in the period of interest for the many crimes they committed (see online Supplemental Material S1). On the other hand, official reactions were relatively lenient (see section *The English and German juvenile justice system*). Against this backdrop, it could be argued that only relatively small official control effects on SAT's causes of crime can be expected among the examined English and German offenders.

H1: Official contact, on average, has no or only a weak impact on personal morals.
H2: Official contact, on average, has no or only a weak impact on deviant peer associations.
H3: Official contact, on average, has no or only a weak impact on detection risk perceptions.

Data and Methods

Samples and Design

The panel studies *Peterborough Adolescent and Young Adult Development Study* (PADS+; Wikström et al., 2012) and *Crime in the modern City* (CrimoC; Boers et al., 2010) provide longitudinal information from students in the cities of Peterborough (England) and Duisburg (Germany) who were 13 years old at the beginning of the projects.[6] The data in both studies stem from questionnaires and administrative registers providing self-reported information on various domains (including a person's temperament, family, neighborhood, and criminal involvement) and official records of the participants' criminal and sanctioning history.

PADS+ achieved its goal of sampling approximately one-third of a Peterborough student cohort by gathering information from 710 youths in its first wave of interviews in 2004. CrimoC's objective to survey an entire Duisburg school cohort, on the other hand, resulted in the collection of valid data from 3,411 juveniles in its first wave in 2002 (i.e., about two-thirds of the student population in Duisburg). Due to high retention efforts, PADS+ was characterized by exceptionally low panel attrition in the follow-up waves, which were first conducted annually and later at longer intervals. Of the 710 students, who responded in the first wave, 693 still participated in wave 5. Besides, 700 participants consented to the collection of their official contact records from the *Police National Computer* system. Unlike PADS+, CrimoC allowed new students to enter the study in follow-up panel waves (e.g., by entering a participating school). Partly due to the differences in design, CrimoC's retention rate was somewhat lower than that of PADS+. However, participation was still satisfactorily high, with more than 3,200 participants taking part in the first five waves. In CrimoC's fourth wave, most students (87.0%; 2,964 of 3,405) also consented to the collection of their official criminal and sanctioning records from the *Erziehungsregister* and the *Bundeszentralregister*.[7]

To guarantee that each individual provides sufficient information for the treatment effect estimation, I defined two conditions for inclusion in the *final analysis samples*. Juveniles were included if they (1) participated in panel waves 3–5 and (2) consented to the collection of their official records. These criteria applied to 690 PADS+ students, who make up most of the original respondents (97.2%; 690 of 710). In CrimoC, the analysis sample consists of 2,117 juveniles, a more reduced subset of the total sample (62.2%; 2,117 of 3,405 participants in wave 4). Fulfilling the conditions less likely, youth with a high risk of criminal behavior and sanctioning experiences were disproportionately excluded from analyses in CrimoC (see online Supplemental Material S3.2). Hence, treatment effect estimates may not be representative for Duisburg's entire juvenile population but may be rather sample-specific.[8]

Measures

The current study distinguishes three periods: *pretreatment* (T1), *treatment* (T2), and *posttreatment* (T3). This breakdown makes it easier to refer to particular periods in PADS+ and CrimoC and helps establish a proper causal time order between the covariates, treatment, and outcomes (Table 2). The T2 treatment measures include information on system contacts in the year 2006 in PADS+ and in the year 2004 (March–December) in CrimoC. The T1 covariates and T3 outcomes (causes of crime), on the other hand, are mainly based on self-reports from questionnaires that were completed typically in the first half of the year before or after the system contact, respectively. Thus, the time between T2 official contact and T3 outcomes ranges between 1 month and a year, depending on the time the contact took place and the time the respondent completed the T3 questionnaire. The study, thus, analyzes the short-term impact of formal control on the causes of crime, which may differ from its long-term consequences as some influences may unfold only over a longer time period.

The following presentation of the measures is restricted to the most crucial concepts. Online Supplemental Materials S3.1 and S3.2 provide additional insights into the measurements. The materials also give information about item nonresponse. Overall, CrimoC was much more affected by item-missingness than PADS+. While only 5.1% (35 of 690) of the PADS+ juveniles had at least one missing value in the covariates, treatment, or outcomes, the same was true for most CrimoC participants (69.2%; 1.465 of 2,117). However, the average missing data proportion per individual was relatively low in both studies (PADS+: <1%; CrimoC: 3.0%) and was dealt with by multiple imputation in the current study (see section *Analytical procedures*).

Outcomes. All outcome variables are generated by computing the mean score of the indicators that measure the respective concept. Cronbach's alpha of the scores varies from 0.76 to 0.87 in PADS+ and 0.86 to 0.93 in CrimoC, indicating that the concepts were measured quite reliably in both studies.

PADS+ operationalized personal morals with three concepts, including personal moral rules (cognitive component) and anticipated emotions of guilt and shame when breaking a particular legal rule (emotive component). CrimoC, on the other hand, collected information about personal moral rules and additionally measured youths' general legal rule acceptance. The measures in PADS+ were explicitly developed to test SAT and thus closely resemble the concept of personal morals as indicated by the theory. CrimoC's two measures, on the other hand, map the cognitive component of morality but do not (or only to a small degree) consider its emotive parts, depicted mainly by the emotions of guilt and shame according to SAT.

Personal moral rules were operationalized very similarly in both projects. Participants assessed how wrong (PADS+) or bad (CrimoC) they thought several delinquent acts were. For the evaluation of five offenses (e.g., burglary, graffiti spraying, robbery, shoplifting, and smashed streetlight),

Table 2. Causal Time Order of Measures.

Phase	Ø-age	Time periods		Measures
		PADS+	**CrimoC**	
T1	14	January–December 2005	January 2003–February 2004	Covariates
T2	15	January–December 2006	March–December 2004	Official contact
T3	16	January–May 2007	January–April 2005	Causes of crime

Note: CrimoC's treatment period (T2) is shorter to take into account that some covariates (e.g., self-reported delinquency in T1) refer to the period from January 2003 to January/February 2004, whereas comparable measures in PADS+ refer only to the whole years (e.g., the entire year 2005). CrimoC = Crime in the modern City; PADS+ = Peterborough Adolescent and Young Adult Development Study.

PADS+ participants had four different response categories available ranging from (0) "not wrong at all" to (3) "very wrong." CrimoC respondents, on the other hand, evaluated seven delinquent acts (assault, burglary, bicycle theft, extortion, shoplifting, theft of a car, vandalism) using categories from (0) "entirely harmless" to (4) "very bad."

Moral emotions of guilt and shame were measured in PADS+ with questions concerning whether one would feel guilty or ashamed (in front of significant others) when committing different deviant or criminal acts (guilt: assault, break a parental rule, cheated on test in school, shoplifting, teased another school/work mate, theft from a car; shame: best friends/parents/teachers found out about shoplifting, best friends/parents/teachers found out about breaking into a car) and, if yes, how much. Response categories for the six items ranged from (0) "No, not at all" to (2) "Yes, very much."

CrimoC, instead, operationalized individuals' *general acceptance of legal rules* by asking about why one should not commit crimes. The ten listed reasons for abiding by the law were: "most of the time you will get caught," "you just shouldn't do that," "it is important to respect the law," "it is important to follow the rules that others should obey, too," "it is worthwhile to have a clear conscience," "you are harming others who are innocent," "it is important to be a good example for others (e.g., children)," "it would be detrimental to me," "the possible sanction is just too high," "delinquency damages the reputation of one's family," "it is important to respect the law," and "it is valuable to have a clear conscience." Participants rated these statements on a scale from (0) "strongly disagree" to (4) "strongly agree."

Association with deviant peers (or *peer deviancy*) was measured in PADS+ by asking juveniles about their friends' involvement in six different deviant acts (e.g., assault/get into fights, get drunk, use drugs, shoplifting, skipping school, and vandalism). They rated the frequency of involvement on a scale ranging from (0) "No, never" to (3) "Yes, very often (every week)." CrimoC, on the other hand, collected information about participants' *deviant peer group association* using six items: "There are other opposing groups.", "We also use violence to pursue our interests.", "We fight with other groups.", "When we show up together, others truly have respect.", "When we're together, we drink a lot of alcohol, too.", and "To have fun, we sometimes do something illegal." The participant assessed how much these statements applied to their peer group using the response categories (0) "disagree" to (4) "totally agree".

Finally, the *perception of detection risk* (or deterrence sensitivity) was operationalized very similarly in both studies by asking the juveniles to estimate the risk of getting caught for various crimes. In PADS+, the detection risk for four offenses (assault, shoplifting, theft from a car, vandalism) was rated using the response categories (0) "No risk at all" to (3) "A very great risk." In CrimoC, participants evaluated the risk of getting caught for seven different crimes (assault, bicycle theft, burglary, extortion, shoplifting, theft of a car, vandalism) on a scale ranging from (0) "very unlikely" to (4) "very likely."

Treatment: Juvenile justice contact, the treatment, differentiates between juveniles who had an official record of at least one criminal activity within period T2 and those who did not. While in PADS+, 37 (out of 690; 5.4%) participants had an official record of contact with the juvenile justice system within T2, the same is true for 88 (out of 2,117; 4.2%) CrimoC participants. Most of the treated respondents generally experienced a rather lenient system contact in both systems (Table 1). Most cases were diverted out of the official system (PADS+: 27 out of 37, 73.0%; CrimoC: 71 out of 88, 80.7%), and only a very few juveniles were imprisoned (PADS+: 1 out of 37, 2.7%; CrimoC: 3 out of 88, 3.4%). Hence, most individuals who were officially recorded for a crime at T2 had their case dismissed with only minimal system contact (typically including a stern talking to by police officers or an official latter by the prosecution), or had their case dismissed but were encouraged to participate voluntarily in a rehabilitation program (England) or on the condition that there was some educational intervention by the youth's parents, school, or employer

(Germany). Although the studied system contact was the first for most treated juveniles, some had had contact before (16.2% or 12.5% had an official contact at T1 in PADS+ or CrimoC, respectively).

Covariates: Following recommendations by experts (e.g., Kainz et al., 2017; Morgan & Winship, 2015), the selection of 52 covariates was based on theories that either explained why juveniles had contact with the juvenile justice system or/and why they varied in the outcomes of interest. All covariates selected in this way either refer to a time within T1 or are time-invariant. They include the baseline (T1) outcomes because these are among the most helpful variables to condition on when identifying causal estimates (Morgan & Winship, 2015; Steiner et al., 2010). In addition to these lagged outcomes, other covariates come from different domains, including a person's criminal and deviant history, personal characteristics (e.g., temperament/character), and social associations (e.g., peers and family). For each domain, multiple indicators were included to diminish selection effect threats (Steiner et al., 2010; for a list of all domains and covariates, see online Supplemental Materials S3.1 and S3.2).

Analytical Procedures

For all outcomes, the estimand of interest is the *Average Treatment Effect on the Treated* (ATT; Morgan & Winship, 2015). The ATT is generally defined as the average difference in two potential outcomes, Y_i^1 and Y_i^0, among individuals who were actually treated ($\mathrm{Tr} = 1$)[9]:

$$\mathrm{ATT} = E[Y_i^1 - Y_i^0 | \mathrm{Tr} = 1] = E[Y_i^1 | \mathrm{Tr} = 1] - E[Y_i^0 | \mathrm{Tr} = 1].^{10} \quad (1)$$

In this study, Y_i^1 refers to a juvenile's outcome (e.g., risk perception) at T3 if she or he had had official contact in T2. Y_i^0, in contrast, refers to the potential outcome at T3 if the same juvenile had not had official contact in T2. Because treated juveniles had official contact in T2 by definition (in the current study: treatment = official contact), Y_i^1 is observed for all of them. It can, therefore, easily be entered into the ATT equation. The information for Y_i^0 is, however, missing among the treated individuals because they did not experience the counterfactual state in which they had no system contact. Therefore, the ATT cannot be computed with observed data alone.

I used matching methods to infer the ATT for the juveniles with official contact from their counterparts with no official contact who had similar values on all key pretreatment characteristics (Stuart, 2010). In particular, I followed four matching steps. The goal of the *first three steps* was to find a *propensity score matching* (Apel & Sweeten, 2010; Rosenbaum & Rubin, 1983) procedure that best balanced the treated and control group in terms of their covariate distributions. A propensity score refers to the probability that an individual is treated. In the first step of the matching procedure, I estimated multiple propensity scores for each individual using the covariates as predictors in three different estimation procedures.[11] Second, I utilized the various computed propensity scores in different matching algorithms.[12] In the third step, I selected the best propensity score matching combination using the so-called balance statistics that indicate how well balanced the treated and untreated groups are in terms of covariate distributions after the respective matching (Kainz et al., 2017). The same propensity score matching procedure does not need to be the best in balancing both the English and German samples. In fact, different matching techniques delivered the best balance for each sample. Therefore, ATT estimates derived by different matching procedures are reported for each sample. This trial and error approach in the first three matching steps is recommended because a good covariate balance diminishes the threat of selection bias due to pretreatment differences in observables (e.g., Kainz et al., 2017; Morgan & Winship, 2015).

Whatever the most successful matching technique, the resulting best-balanced samples were finally used in the fourth step in regression analyses to estimate the ATTs and their uncertainty. In

particular, normal linear regressions were applied using the measures of personal morals, deviant peer associations, and risk perceptions one after the other as dependent variables. In each case, the regression model included as predictors the treatment indicator, the lagged outcome variable, and their interaction term.[13] The matched nature of the sample was taken into consideration using weights—the final product of the first three analytical steps.

Because the final samples were affected by item nonresponse, all analytical steps (e.g., propensity score estimation, matching, and outcome analyses) were applied to multiple imputed data sets.[14] As recommended by Penning de Vries and Groenwold (2017), propensity score estimation, matching, the generation of weights, and outcome analyses were conducted for each imputed data set. The imputed information was finally combined by merging the vectors of all ATT simulations.

Each step in the analytical procedure was, furthermore, conducted using different combinations of multiple imputation, propensity score estimation, matching, and outcome methods to check the robustness of the ATT estimates to different method combinations (Young & Holsteen, 2016). Overall, 48 method combinations in PADS+ and 60 in CrimoC were included as promising candidates in model dependence assessments. Promising candidates are those method combinations that successfully balanced the treatment and control groups regarding covariate distributions. To assess model sensitivity, I computed the ATT median point estimate of each candidate model and plotted their distribution using density plots.

Results

This section first reports the covariate balance before and after adjustment. In doing so, it addresses how well each selected matching scheme balanced the groups of individuals with and without official contact and thereby diminished the threat of selection bias. In the second step, the section presents the ATTs computed with the help of the adjusted samples. Finally, the model dependence of the ATT estimates is assessed.

Pretreatment Covariate Balance

As recommended by experts (Kainz et al., 2017), I assess the covariate balance using the balance statistics standardized bias (SB) and variance ratios (VRs). SB is the difference in covariate means between the treated and control group divided by the treated individuals' *SD*. VRs of continuous covariates are computed by taking the ratio of variances observed in the treatment and control group for the covariate at hand. The literature identifies SB values below 0.1 (0.2, 0.25; the literature is not settled on a threshold) and VRs between 0.5 and 2 as an indicator of a reasonable balance in a covariate's distribution across the treatment and control group (Harder et al., 2010; Kainz et al., 2017).[15] Due to limitations of space, Tables 3 and 4 report balance statistics only for the lagged outcomes. Online Supplemental Material S4 includes balance statistics for the rest of the covariates.

English sample (PADS+). Before matching, treated and untreated English juveniles differed substantially concerning key pretreatment characteristics (Table 3). Most (42 of the 52) covariates exhibited SBs larger than 0.1, thus indicating an imbalance between the groups. Many (29) covariates still exceeded a less stringent threshold of 0.2. The mean (0.16) and the median (0.10) of the absolute SB across variables were also beyond the threshold. All lagged outcome variables had absolute SB values of 0.4 or larger, with the highest standardized difference in the peer deviancy measure. Its difference of 0.71 implies that individuals with official contact reported much more peer deviancy at T1 than individuals without official contact. Contrary to SB statistics, VRs of continuous covariates point to only a few substantial differences in covariate distributions across the treated and untreated groups in PADS+. Most continuous covariates (13 out of 16) had VRs of well below 2,

Table 3. Covariate Balance Satistics for the English Sample (Short Version).

	Unadjusted Sample		Adjusted Sample	
English sample (PADS+)	**SB**	**VR**	**SB**	**VR**
Covariates—Lagged (T1) Outcomes				
Personal moral rules	−.53	2.04	−.05	1.10
Moral shame	−.55	1.52	−.02	1.27
Moral guilt	−.50	1.66	.05	1.40
Deviant peers	.71	1.32	−.17	1.38
Detection risk perceptions	−.42	1.86	−.01	1.68
Global Covariate Balance Statistics				
Mean (absolute)	.16	1.63	.05	1.28
Median (absolute)	.10	1.40	.04	1.28
Maximum (absolute)	1.02	3.47	.32	1.74

Note: PADS+ = Peterborough Adolescent and Young Adult Development Study; SB = standardized bias; VR = variance ratio.

and the mean and median of the VRs across covariates were only 1.63 and 1.40, respectively. Overall, balance statistics indicate that groups were quite imbalanced before matching and that this was especially true for the lagged outcome variables.

According to the balance statistics, *optimal matching* with a 1:3 ratio without replacement on the linear propensity score estimated via Bayesian additive regression trees (Chipman et al., 2010) best balanced the English treatment and control groups. Optimal matching is a procedure that minimizes a global distance measure by matching control units to treated ones (Hansen, 2004). The matched English sample included an adjusted number of 148 (37 treated and 111 control) juveniles. Optimal matching significantly improved the balance across the treated and control groups (Table 3). After the procedure, 17 covariates exceeded the SB threshold of 0.1, whereas the same was true for only one variable considering the less stringent 0.2 threshold. The mean and median absolute SB across covariates also diminished (from 0.16 to 0.10) to values of 0.05 and 0.04. Juveniles with official contact, however, were still slightly more involved with legal actors before their contact, had lower self-control capabilities, perceived the risks of consequences (when committing crime) as lower, had more disadvantaged families, reported more informal social control in their neighborhood, and were less successful and socially integrated into school than juveniles without official contact (see online Supplemental Material S4). In terms of the lagged outcome variables, treated and untreated English juveniles were balanced satisfactorily after matching, except for peer deviancy. Juveniles with official contact reported less (SB = −0.17) peer deviancy than matched youths without official contact. VRs were all well below the threshold of 2 after matching. Overall, treated and untreated groups are much more similar as a result of the optimal matching procedure. Hence, we are much closer to comparing "apples to apples" in the forthcoming analyses, as Morris and Piquero (2013, p. 848) put it. Accordingly, it is less likely that differences in pretreatment characteristics confound the ATT estimates derived with the matched sample. To further reduce the risk that the remaining imbalances bias the treatment effect estimates, the regression models used for ATT estimation include the respective lagged outcome variable as a predictor (see section *Analytical procedures*).

German sample (CrimoC). In the German sample, covariate imbalance before adjustment was considerably lower than in the English sample (Table 4). Only half of the covariates (28 out of 52) exceeded the 0.1 SB threshold (10 covariates the threshold of 0.2). Across all variables, the mean and median absolute SBs were just 0.07 and 0.04. However, all lagged outcome variables exceeded the 0.1

Table 4. Covariate Balance Statistics for the German Sample (Short Version).

	Unadjusted Sample		Adjusted Sample	
German sample (CrimoC)	**SB**	**VR**	**SB**	**VR**
Covariates—Lagged (T1) Outcomes				
Personal moral rules	−.12	1.23	.00	1.06
General legal norm acceptance	−.35	1.46	−.03	1.12
Deviant peer group association	.37	1.35	.03	1.06
Detection risk perceptions	−.21	1.04	.00	1.12
Global Covariate Balance Statistics				
Mean (absolute)	.07	1.23	.01	1.13
Median (absolute)	.04	1.24	.00	1.10
Maximum (absolute)	.37	1.46	.03	1.51

Note: CrimoC = Crime in the modern City; SB = standardized bias; VR = variance ratio.

threshold, and deviant peer group association, as well as general legal norm acceptance, were among the most imbalanced of all variables, with values of 0.37 and −0.35, respectively. Juveniles who had official contact at T2 reported that at T1, they were more strongly involved in deviant peer groups and felt less bound to legal norms than their counterparts without official contact. VRs, in contrast, were already all well below the threshold of 2, indicating no substantial differences in the variances of the continuous covariates across the treated and the untreated groups. In a nutshell, covariate distributional differences, although less substantial than in the English sample, were still large enough to potentially bias treatment effect estimates.

The balance statistics indicate that the method that best balanced the German treatment and control groups was *weighting by the odds* on the Covariate Balancing Propensity Score (Imai & Ratkovic, 2014). Weighting by the odds puts more weight on individuals from the control group who are more similar to treated persons on the propensity score than on less similar control units to weight the whole control group up to the treatment group (Harder et al., 2010). The weighted sample includes an (efficient) adjusted number of 300.6 (88 treated and 212.6 control) juveniles. Weighting by the odds diminished the distributional imbalance in observed covariates across the board (see Table 4; online Supplemental Material S4): Not a single covariate was imbalanced after weighting. The mean of the absolute SB across variables is very close to null (0.01; median: 0.00), and the VRs' value of 1.13 is not far from a perfect variance balance across treated and untreated participants. For the German sample, the weighting procedure likely prevents potential selection bias due to observed covariates. To use Morris and Piquero's (2013) language, it appears we are comparing the same varieties of apples in the weighted treated and control groups. Consequently, the likelihood of confounding by pretreatment characteristics is vastly diminished.

Effects of System Contact on the Causes of Crime

The effects of official contact on the causes of crime (estimated with equation 1) were mostly relatively small in both studies (see gray-shaded areas or black solid lines and dots in Figure 2). First, official contact hardly affected personal morals, with most effect estimates being not significantly different from null. For example, young offenders in PADS+ changed *personal moral rules* on average only by 0.01 [89% *CI*[16]: −0.08, 0.11] and in CrimoC by 0.08 [89% *CI*: −0.08, 0.23] because of official contact. Additionally, formal contact seemed not to or only weakly influence the *moral shame* of English offenders (0.00 [89% *CI*: −0.08, 0.07]) and the *general legal norm*

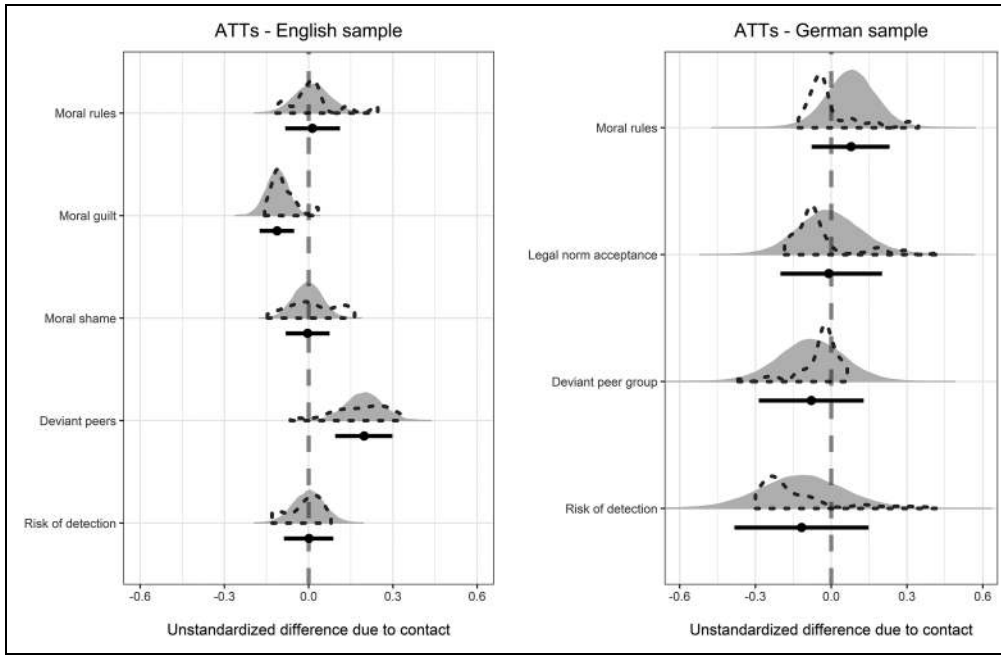

Figure 2. Average Treatment Effect on the Treated (ATT) estimates.
Note: Gray-shaded area = Distribution of ATT simulations of the best-balancing model; Black dots = Medians of ATT simulations of the best-balancing model; Black lines = 89% *CIs* of the best-balancing model; Dotted black lines = Distribution of medians of the ATT simulations of all candidate models.

acceptance of German offenders (−0.01 [89%-*CI*: −0.20, 0.20]). However, as an exception, the ATTs suggest that *moral guilt* among English offenders may have been affected more substantially by official contact. The effect probably ranged from modestly decreasing to rather unsubstantial (−0.11 [89% *CI*: −0.17, −0.05]). Whereas juveniles with official contact anticipated that they would, on average, not have felt guilty or would have felt only a little guilty after committing a crime ($E[Y^1] = 0.85$), estimates indicate that they would have felt slightly more guilty on average if they had not had official contact ($E[Y^0] = 0.96$ [89% *CI*: 0.90, 1.02]).

Second, ATT estimates suggest that the *deviant peer group association* of German offenders was not substantially affected by system contact (−0.08 [89% *CI*: −0.29, 0.13]). In contrast, English juveniles reported more *peer deviancy* in the following year on average. The ATT indicates that treated English juveniles had around 0.20 [89% *CI*: 0.09, and 0.30] higher peer deviancy scores after official contact with the justice system in T2 compared to a (hypothetical) situation in which they would not have had official contact. The model implies that the treated youths would have reported a deviant peer score of around 1.34 [89% *CI*: 1.24, 1.44] on average if they had not had official contact in the previous year, while their observed average score was 1.54.

Finally, according to the treatment estimates, official contact appears to have had no substantial effect on juveniles' detection risk perceptions. The effect estimates for treated German participants are, however, relatively imprecise. They allow for both small increases and moderate decreases in detection risk perception due to official contact (−0.12 [89% *CI*: −0.38, 0.15]). Therefore, we cannot learn much from this estimate. In contrast, the ATT for the English sample indicates that the official control effect on detection risk perception is at best relatively small (0.00 [89% *CI:* −0.09, 0.09]).

Assessment of Model Robustness

The robustness of the ATT estimates to changes in the modeling procedure was assessed by density plots (see dotted lines in Figure 2). They imply that the treatment estimates are generally more model-sensitive in the English than in the German sample.[17]

Whereas ATT median point estimates for the English youths were relatively insensitive to alternative modeling procedures for moral rules, moral guilt, and detection risk perception, the effect estimates on moral shame and peer deviancy were less robust. For moral shame, point effect estimates vary from small increases to small decreases depending on the methodological approach used. For peer deviancy, density plots imply that effect sizes may vary from relatively unsubstantial (although positive) to moderate increases. Consequently, the latter two outcomes must be discussed with more caution than the others.

For the German sample, the medians of the ATT estimates are all quite robust to changes in the methodological approach. Especially for moral rules, legal norm acceptance, and deviant peer group association, different methodological approaches nearly all lead to very similar results. The outcome that is most sensitive to the modeling procedure is the detection risk perception. The different methodological approaches produced estimates indicating that the effects on risk perceptions ranged from relatively unsubstantial to weakly decreasing. Against this background, ATTs of risk detection must be interpreted with caution for the German sample.

Discussion

Why do criminal justice contacts often have no crime-preventative impact? To shed light on this question, the current study examined the influence of system contact on factors that may mediate official contact effects on reoffending. Unlike previous empirical studies, which mostly tested either deterrence or labeling theory, the current study defined SAT's causes of crime as potential mediators. SAT was chosen as the theoretical framework primarily for the following reasons. First, SAT provides terminology that clearly distinguishes between factors that directly cause crime and more distal factors. Second, using SAT as a theoretical framework allows for simultaneous integration of assumptions about contact-induced mechanisms by both deterrence and labeling theory. The causes of crime identified by SAT and selected for analysis were personal morals, deviant peer associations, and risk perceptions. The analysis explored whether contact with the English and German juvenile justice system influenced these causes of crime. The results supplement the sparse body of research outside the United States that had previously studied the effects of system contact on antecedents of crime.

The ATT estimates suggest that the most lenient system contacts in this study had relatively minor effects on most causes of crime. In the German sample, neither the young offenders' morals nor their deviant peer group association or detection risk perceptions were significantly—or substantially—affected by official contact. In the English sample, three of the five effects were statistically insignificant. The significant estimates suggest that contact had a weak guilt-reducing effect and moderately amplified peer deviancy among the English juveniles (with the latter finding challenged by some sensitivity analyses producing smaller effect estimates). Overall, the results indicate—in line with Hypotheses H1 to H3—that official controls did not (CrimoC) or only weakly (PADS+) trigger crime-relevant processes. Although not analytically testable due to different measures, the results seem furthermore to suggest that effects were more likely among English than German youths. These more substantial findings in the English sample indicate that system contact triggers criminogenic rather than crime-preventative mechanisms. This missing or adverse impact of official control may explain why most previous studies (including one using the data at hand; Boers et al., 2022) found no crime-preventative effects. The results, furthermore, indicate that the components of

morality may be influenced differentially by a system contact. While the cognitive part (moral rules) was unaffected in both samples, the findings suggest that system contact may have decreased feelings of guilt. Future research should try to replicate these differential effects.

How can it be explained that the current study finds, in contrast to previous research, mostly insignificant effects on the examined intermediate factors and that this is especially true for the German study? The first explanation for these findings may be a low certainty and lenient nature of the system contacts in this study compared to previous studies. The studied juveniles mainly were not detected when they committed a crime, and if they were, were they mostly diverted out of the formal system. Prior studies, in contrast, were primarily based on the U.S. samples, in which the juveniles studied were likely to have been detected with higher certainty and handled more harshly on average due to the more proactive and punitive criminal justice system in the United States (Howell et al., 2013; Huizinga et al., 2003; Snyder & Sickmund, 1999).[18] Deterrence and labeling theory suggests that the more intrusive and certain (consistent) official interventions are, the stronger their effect should be. This assumption is backed up by past empirical findings indicating that the certainty and severity of police contact and sanctions indeed play a crucial role in influencing young offenders' future delinquency and causes of crime (e.g., Anwar & Loughran, 2011; McAra & McVie, 2007; Wiley et al., 2013). The mostly weak control effects found in this study, hence, do not seem to differ substantially from the theories' predictions, as the interventions were lenient and uncertain for most of the treated offenders in this study. Furthermore, the more substantial effects in England may be explained by the fact that the English system reacted more proactively and intensively to crime than the German one.

The second and related explanation for the minor effects of system contact may be that the causes of crime were probably already quite solidified before the contact and, therefore, less malleable. Due to the low risk of detection, juveniles probably already had a relatively high number of informal deterrence experiences before the formal intervention. Through these experiences, they had already learned how likely it is to be caught and how (informal) others react to their misbehavior. These learning processes led over time to quite consolidated morals, risk perceptions, and peer associations. Labeling and deterrence theory indicate that rare and non-intensive formal interventions should have a relatively weak influence on more experienced offenders. As police in England acted more proactive than German police in the studied time period, English offenders may have been less experienced with crime and sanctioning when they had their system contact. This inexperience with formal and informal reactions may explain why official contact was more influential among English youth.

Limitations and Perspectives

This study has some limitations that future research should resolve. First, it cannot test the two arguments that official contact influences the causes of crime differently (a) depending on its nature and certainty and (b) depending on the offender's previous criminal and sanctioning history. To test the argument that more intense interventions should have larger effects, future examinations must differentiate between various levels of system contact (e.g., diversion vs. noncustodial sanctions vs. imprisonment). However, to conduct a more differentiated investigation of this kind, large samples are needed that typically have not been available in existing studies to date (including the current one). For this reason, only a few empirical studies (e.g., McAra & McVie, 2007) have so far distinguished between various intervention levels, let alone conducted analyses of mediating factors and mechanisms to explore how these factors are affected by different sanctions. To test for the effect of the certainty of punishment more directly, future studies should follow the growing deterrence literature on Bayesian updating (e.g., Anwar & Loughran, 2011). This literature typically operationalizes the treatment variable as the ratio of arrests per crime. It thereby directly

explores the impact of the strength of a deterrence signal within a particular period on subsequent risk perceptions. A follow-up study with the present data will rely on such a crime-arrest ratio to study the Bayesian updating predictions more directly. Finally, to test the second argument that contact effects depend on the treated person's criminal and sanctioning history, analyses of subgroups divided by their number of prior crime-related experiences are needed. So far, existing research supports the argument that contact effects are smaller among more experienced offenders (e.g., Chiricos et al., 2007; Ward et al., 2014). However, the few studies that investigated this hypothesis analyzed official contact effects on reoffending. Future research should also study the differential impact of formal interventions on the direct antecedents of crime.

The second limitation of the current study is that it—like all observational designs—is not immune to selection bias arising from potentially unobserved confounders. Having matched on a relatively large number of important observed covariates, this analysis and other more recent studies should, however, be able to account for most of the crucial differences in pretreatment characteristics between treated and untreated individuals. Future research should, nevertheless, be conducted with the explicit goal of studying the impact of criminal justice contact on the causes of crime. It should therefore aim to measure and balance all the covariates that may theoretically confound the effect estimates.

Third, like some previous research (e.g., Wiley et al., 2013), the current study measures the causes of crime by combining items that tap into various forms of deviant and criminal acts. The construction of such global measures, however, is only valid under the assumption that official contact influences various forms of behavior independent of the type of criminal act one was apprehended for. This assumption may be problematic as SAT and other theories (e.g., rational choice approaches) assume that committing a crime and learning from its consequences (e.g., formal contact) are probably action-specific processes (e.g., Anwar & Loughran, 2011; Wikström et al., 2011). Some existing research (including this study) may thus underestimate the effects of official intervention for the specific sanctioned behavior. Therefore, future research should investigate how formal contact for a particular type of criminal activity may influence action-specific morals, peer activities, and risk perceptions (and how this differs from its effect on more global measures). To do this, larger sample sizes are needed that provide the power to analyze offense-specific effects.

Fourth, although the current study takes a cross-national perspective, it cannot formally test the differences in effect sizes between the English and German samples and how they relate to previous U.S.-based studies. That is because the measures in the two studies differ, and no U.S. data were available. The current study's findings and interpretations are thus explorative in nature. They should trigger future research that is already set up cross-nationally from the beginning to explore differential system effects more directly. So far, although called for by experts (e.g., Bernburg, 2019), cross-national designs were applied only very scarcely in the research of sanctioning effects (for an exception, based on samples that were initially not designed for comparison, see Huizinga et al., 2003). However, large-scale cross-national designs should supplement studies that delve deep into one system by showing how various criminal justice systems have a differential impact on apprehended individuals. At best, such studies are also able to consider treatment heterogeneity in more detail by exploring whether differential effects across different settings are due to the prevalence of particular system responses (e.g., one system reacts more with diversion than others) or due to how the same interventions types are implemented across settings (e.g., systems vary in their diversions practices).

Fifth, although the current study is the first one using SAT as a starting framework to investigate official contact effects on potential intervening factors, it does not test SAT's full theoretical model (Figure 1). It focuses on the impact of official contact on specific causes of crime, ignoring the subsequent effect of these causes of crime on delinquency. A structural equation model (SEM) would have been able to study the full model (see Wiley et al., 2013), which could

not be accurately estimated with the data at hand. This is because, according to SAT, the causes of crime interact in a relatively complex way to explain delinquency (Wikström et al., 2012). The study of such interaction effects is only possible with a rather large number of observations. However, this number is typically relatively small in studies based on samples from general populations, as these include only a few juveniles with official contact in a specific period. As this is also true of the current study, it does not rely on SEM. Nevertheless, I believe that the analysis of only a part of the full model is still an informative approach to study the research question at hand, since only when the causes of crime are affected by official contact can they ultimately alter delinquency.

Finally, although, in my view, SAT constitutes a promising framework for a general theory of sanctions, other approaches may be similarly helpful to study official contact effects in their complexity. All general theories of crime may be fruitful starting points for a general theory of sanctions, as they all specify the factors that should be directly causally relevant for offending. Although many often-tested general theories of crime exist, they have—at least to my knowledge—so far not been applied to investigate sanctioning effects empirically. Instead, most studies have tested classic sanctioning theories such as deterrence or labeling theory (e.g., Anwar & Loughran, 2011; Bernburg et al., 2006; Wiley et al., 2013). Despite the value of testing these theories, I agree with Piquero and colleagues (2011) that it may be helpful for future researchers to move beyond the testing of single sanctioning theories. More general frameworks of investigation, as applied here, have the advantage of being able to integrate and study mechanisms postulated by theories as diverse as deterrence and labeling theory.

Conclusion

The findings in the current study imply that official contact significantly affected only a few of the causes of crime in the English juvenile justice system and had basically no impact in the German system. These findings contrast with results of previous U.S. studies, which generally found more substantial control effects on the examined causes of crime. Most of these operated in directions that amplified rather than diminished reoffending. Against this backdrop, a less proactive and punitive control style—as is used in most European countries—may be, on average, the better approach, if not to diminish juvenile delinquency, then at least not to exacerbate it. Diversion, in particular, might be seen as a mechanism that by construction does not trigger strong processes, be they crime-amplifying or crime-preventing. Given that the current state of research indicates that traditional sanctions rather amplify than prevent future crime, diversion may be seen as a more efficient and cost-saving way to handle nonserious juvenile offenses than traditional criminal justice processing (see Petrosino et al., 2014). Juvenile justice systems may make diversion practices even more fruitful by implementing family focused programs that help parents communicate and enforce rules and keep their children away from deviant peers. Research suggests that such family oriented diversion programs may be particularly efficient in reducing delinquency (Schwalbe et al., 2012). This efficiency may be explained by the fact that moral education and parental monitoring are directly related to factors (personal morals, deviant peer associations) that SAT deems as causes of crime (Wikström & Treiber, 2016a).

Acknowledgments

First, I want to thank Björn Huss, Jost Reinecke, Marcus Schaerff, and Hannah Weishaupt for the time they spent discussing earlier drafts of this paper. Additionally, I am indebted to Kevin Wölk for his tireless checking of my coding. I would also like to thank Klaus Boers and Per-Olof Wikström for providing valuable input and the means (e.g., the workspace and data) to work on this project. Finally, I am grateful to the anonymous reviewers for their helpful comments on this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Deutsche Forschungsgemeinschaft (grant number BO 1234/9-1).

ORCID iD

Florian Kaiser https://orcid.org/0000-0002-9686-7738

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Unfortunately, no systematic review of the SAT literature which was published after 2015 exists. My own (probably biased) assessment of the more recent literature is that it produces evidence largely in favor of SAT and that strong research designs produce supportive evidence even more likely (e.g., Wikström et al., 2018).
2. Recent SAT publications focus on developmental aspects (e.g., Wikström, 2020) or crime prevention in general (Wikström & Treiber, 2016a). Accordingly, it can be expected that the theory will be applied more frequently to criminal justice issues in the future.
3. Schulte (2019) used the German data of the current study to examine how the severity of system contact affected some intermediate factors.
4. Some contemporary versions of rational choice theory try to consider why humans often act in a more habitual or automatic manner (e.g., Kroneberg et al., 2010). These theories, however, have so far not been applied to sanctioning research.
5. Although deterrence research (like SAT) acknowledges that the deterrence process is fear-based (see Paternoster, 2010), studies (including the current one) typically analyze risk perceptions. Future research should embrace deterrence as emotion-based process by analyzing how the fear of consequences affects criminal activity and how this fear is shaped by official contact.
6. Online Supplemental Material S2 provides a brief comparison of both cities.
7. For more information on PADS+ and CrimoC, see www.cac.crim.cam.ac.uk/research/padspres and www.crimoc.org or the publications Wikström et al. (2012) and Boers et al. (2010), respectively.
8. It is difficult to speculate on the precise implications of this systematic exclusion of high-risk youths. On the hand, it may diminish treatment effects, as high-risk youth may be exposed to more severe interventions that are deemed to have stronger effects on average. On the other hand, the excluded juveniles may be disproportionately offenders who have had a relatively high number of (informal) deterrence experiences before the official contact. Hence their causes of crime may be quite consolidated and consequently less malleable.
9. From a policy perspective, it seems reasonable to narrow the treatment effect to treated individuals because they are the ones in danger of being apprehended and sanctioned.
10. E[.] refers to the expectation operator from probability theory. In this article, the expectations are averages of particular quantities.
11. I included 35 and 52 covariates as predictors in PADS+ and CrimoC, respectively. For the covariates included in propensity score estimation, see online Supplemental Materials S3.1 and S3.2. Propensity scores were estimated with the following three procedures: Bayesian logistic models (McElreath, 2016), Bayesian additive regression trees (Chipman et al., 2010), and covariate balancing propensity score

estimation (Imai & Ratkovic, 2014). All computations were conducted in *R* (version 3.5.2). For a list of all R packages used, see online Supplemental Material S6.

12. The applied matching algorithms are (Stuart, 2010): (1–5) nearest neighbor matching with replacement, a caliper of 0.25, and ratios of 1:1–1:5, (6–11) optimal matching with ratios 1:1–1:5, and (12) weighting by the odds.
13. For robustness checks, I additionally ran models including (1) only the treatment as independent variable (weighted mean differences) or (2) a relatively rich set of predictors.
14. Multiple imputation embraces the uncertainty in the ATT estimation that is due to missing data by predicting missing values multiple times. I predicted the missing values in the current study for each variable with predictive mean matching within *fully conditional specification* (van Buuren, 2018). In doing so, I produced 70 imputed data sets for CrimoC, but only 12 for PADS+. Because PADS+ was less affected by item nonresponse than CrimoC, fewer imputations should suffice to generate reproducible results (van Buuren, 2018). In addition to predictive mean matching, I also applied other imputation procedures (e.g., random forests). These sensitivity analyses show that the ATT estimates are relatively robust to the type of imputation technique used (see online Supplemental Material S5).
15. VRs were standardized so that they are always larger than 1. Consequently, ratios above 2 indicate covariate imbalance. For categorical covariates, raw proportional differences are used as balance statistics, although no thresholds exist for them so far (Kainz et al., 2017).
16. I report 89% *CIs* instead of the usual 95% *CIs* to highlight that the classic choice of a 95% *CI* over any other interval is arbitrary, and that the end of a *CI* should not be interpreted as particularly important but just as a description of how many simulations (e.g., 89%) lie within a particular range (see McElreath, 2016). I also abstain from reporting *p*-values as these encourage binary thinking (significant vs. not significant).
17. For more information on the model robustness of the ATTs, see online Supplemental Material S5.
18. A cross-national study of the U.S. and German system, for example, shows that the percentage of officially arrested juveniles was much higher in the United States than in Germany and that the arrested U.S. juveniles were dealt with more harshly by the juvenile justice system than their German counterparts (Huizinga et al., 2003).

References

Ageton, S. S., & Elliott, D. S. (1974). The effects of legal processing on delinquent orientations. *Social Problems*, *22*(1), 87–100. https://doi.org/10.2307/799569

Albrecht, H.-J. (2010). *Kriminologie* (4th ed.). Beck.

Anwar, S., & Loughran, T. A. (2011). Testing a Bayesian learning theory of deterrence among serious juvenile offenders. *Criminology; An Interdisciplinary Journal*, *49*(3), 667–698. https://doi.org/10.1111/j.1745-9125.2011.00233.x

Apel, R. J., & Sweeten, G. (2010). Propensity score matching in criminology and criminal justice. In A. R. Piquero, & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 543–562). Springer.

Barrick, K. (2014). A review of prior tests of labeling theory. In D. P. Farrington, & J. Murray (Eds.), *Labeling theory: empirical tests* (pp. 89–112). Transaction Publishers.

Bateman, T. (2017). *The state of youth justice 2017: An overview of trends and developments*. National Association for Youth Justice.

Beier, H. (2018). Situational peer effects on adolescents' alcohol consumption: The moderating role of supervision, activity structure, and personal moral rules. *Deviant Behavior*, *39*(3), 363–379. https://doi.org/10.1080/01639625.2017.1286173

Bernburg, J. G. (2019). Labeling theory. In M. D. Krohn, N. Hendrix, G. P. Hall, & A. J. Lizotte (Eds.), *Handbook on crime and deviance* (pp. 179–196). Springer.

Bernburg, J. G., Krohn, M. D., & Rivera, C. (2006). Official labeling, criminal embeddedness, and subsequent delinquency: A longitudinal test of labeling theory. *Journal of Research in Crime and Delinquency*, *43*(1), 67–88. https://doi.org/10.1177/0022427805280068

Boers, K., Reinecke, J., Seddig, D., & Mariotti, L. (2010). Explaining the development of adolescent violent delinquency. *European Journal of Criminology*, *7*(6), 499–520. https://doi.org/10.1177/1477370810376572

Boers, K., Kaiser, F., Schaerff, M., & Wikström, P.-O. H. (2022). The impact of formal controls: A comparative study between an English and a German city. *[Manuscript under Review]*. 1–46.

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, *4*(1), 266–298. https://doi.org/10.1214/09-AOAS285

Chiricos, T., Barrick, K., Bales, W., & Bontrager, S. (2007). The labeling of convicted felons and its consequences for recidivism. *Criminology; An Interdisciplinary Journal*, *45*(3), 547–581. https://doi.org/10.1111/j.1745-9125.2007.00089.x

Dünkel, F., & Heinz, W. (2017). Germany. In S. Decker, & N. Marteache (Eds.), *International handbook of juvenile justice* (2nd ed., pp. 305–326). Springer.

Farrington, D. P., & Murray, J. (2014). Empirical tests of labeling theory in criminology. In D. P. Farrington, & J. Murray (Eds.), *Labeling theory: empirical tests* (pp. 1–9). Transaction Publishers.

Flanagan, R. (2007). *The review of policing: interim report*. Home Office.

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, *99*(467), 609–618. https://doi.org/10.1198/016214504000000647

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, *15*(3), 234–249. https://doi.org/10.1037/a0019623

Home Office, & Youth Justice Board (2002). *Final warning scheme: Guidance for the police and youth offending teams*.

Howell, J. C., Feld, B. C., Mears, D. P., Farrington, D. P., Loeber, R., & Petechuk, D. (2013). *Bulletin 5: young offenders and an effective response in the juvenile and adult justice systems: what happens, what should happen, and what we need to know*. U.S. Department of Justice.

Huizinga, D., & Henry, K. L. (2008). The effect of arrest and justice system sanctions on subsequent behavior: findings from longitudinal and other studies. In A. M. Liberman (Ed.), *The long view of crime: A synthesis of longitudinal research* (pp. 220–254). Springer.

Huizinga, D., Schumann, K., Ehret, B., & Elliott, A. (2003). *The effect of juvenile justice system processing on subsequent delinquent and criminal behavior: A cross-national study [Final report to The National Institute of Justice]*. University of Bremen; University of Colorado.

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 243–263. https://doi.org/10.1111/rssb.12027

Johnson, L. M., Simons, R. L., & Conger, R. D. (2004). Criminal justice system involvement and continuity of youth crime: A longitudinal analysis. *Youth & Society*, *36*(1), 3–29. https://doi.org/10.1177/0044118X03260323

Kainz, K., Greifer, N., Givens, A., Swietek, K., Lombardi, B. M., Zietz, S., & Kohn, J. L. (2017). Improving causal inference: recommendations for covariate selection and balance in propensity score methods. *Journal of the Society for Social Work and Research*, *8*(2), 279–303. https://doi.org/10.1086/691464

Kaiser, F. (2021). The influence of differential peer associations on delinquency varies depending on the combination of unstructured socializing and personal morals. *Kriminologie - Das Online-Journal | Criminology - The Online Journal*, *3*(1), 72–99. https://doi.org/10.18716/ojs/krimoj/2021.1.5

Kaplan, H. B., & Johnson, R. J. (1991). Negative social sanctions and juvenile delinquency: effects of labeling in a model of deviant behavior. *Social Science Quarterly*, *72*(1), 98–122.

Kleck, G., & Sever, B. (2017). *Punishment and crime: The limits of punitive crime control*. Routledge. https://doi.org/10.4324/9781315142258

Kreager, D. A., & Matsueda, R. L. (2014). Bayesian Updating and crime. In G. J. N. Bruinsma, & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 116–125). Springer. https://doi.org/10.1007/978-1-4614-5690-2_397

Krohn, M. D., & Lopes, G. (2015). Labeling theory. In M. D. Krohn, & J. Lane (Eds.), *The handbook of juvenile delinquency and juvenile justice* (pp. 312–330). John Wiley & Sons, Inc.

Krohn, M. D., Lopes, G., & Ward, J. T. (2014). Effects of official intervention on later offending in the Rochester youth development study. In D. P. Farrington, & J. Murray (Eds.), *Labeling theory: empirical tests* (pp. 179–208). Transaction Publishers.

Kroneberg, C., Heintze, I., & Mehlkop, G. (2010). The interplay of moral norms and instrumental incentives in crime causation. *Criminology; An Interdisciplinary Journal*, *48*(1), 259–294. https://doi.org/10.1111/j.1745-9125.2010.00187.x

Lochner, L. (2007). Individual perceptions of the criminal justice system. *American Economic Review*, *97*(1), 444–460. https://doi.org/10.1257/aer.97.1.444

Matsueda, R. L., Kreager, D. A., & Huizinga, D. (2006). Deterring delinquents: A rational choice model of theft and violence. *American Sociological Review*, *71*(1), 95–122. https://doi.org/10.1177/000312240607100105

McAra, L., & McVie, S. (2007). Youth justice? The impact of system contact on patterns of desistance from offending. *European Journal of Criminology*, *4*(3), 315–345. https://doi.org/10.1177/1477370807077186

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press.

Ministry of Justice (2010). *Youth crime: Young people aged 10-17 receiving their first reprimand, warning or conviction 2000-01 to 2009-10*.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: methods and principles for social research* (2nd ed.). Cambridge University Press.

Morris, R. G., & Piquero, A. R. (2013). For whom do sanctions deter and label? *Justice Quarterly*, *30*(5), 837–868. https://doi.org/10.1080/07418825.2011.633543

Murray, J., Blokland, A. A. J., Farrington, D. P., & Theobald, D. (2014). Long-term effects of conviction and incarceration on men in the cambridge study in delinquent development. In D. P. Farrington, & J. Murray (Eds.), *Labeling theory: empirical tests* (Vol. 18, pp. 209–235). TRansaction.

Office for Criminal Justice Reform (2004). *Cutting crime, delivering justice: A strategic plan for criminal justice 2004-08*.

Paternoster, R. (2010). How much do we really know about criminal deterrence? *Journal of Criminal Law and Criminology*, *100*(3), 765–824.

Paternoster, R. (2018). Perceptual deterrence theory. In D. S. Nagin, F. T. Cullen, & C. L. Jonson (Eds.), *Deterrence, choice, and crime* (pp. 81–106). Routledge.

Paternoster, R., & Iovanni, L. (1989). The labeling perspective and delinquency: An elaboration of the theory and an assessment of the evidence. *Justice Quarterly*, *6*(5), 359–394. https://doi.org/10.1080/07418828900090261

Pauwels, L. J. R., Svensson, R., & Hirtenlehner, H. (2018). Testing situational action theory: A narrative review of studies published between 2006 and 2015. *European Journal of Criminology*, *15*(1), 32–55. https://doi.org/10.1177/1477370817732185

Penning de Vries, B. B. L., & & Groenwold, R. H. H. (2017). A comparison of two approaches to implementing propensity score methods following multiple imputation. *Epidemiology, Biostatistics and Public Health*, *14*(4), 1–21. https://doi.org/10.2427/12630

Petrosino, A., Turpin-Petrosino, C., & Guckenburg, S. (2014). The impact of juvenile system processing on delinquency. In D. P. Farrington, & J. Murray (Eds.), *Labeling theory: empirical tests* (pp. 113–147). Transaction Publishers.

Piquero, A. R., Paternoster, R., Pogarsky, G., & Loughran, T. (2011). Elaborating the individual difference component in deterrence theory. *Annual Review of Law and Social Science*, *7*(1), 335–360. https://doi.org/10.1146/annurev-lawsocsci-102510-105404

Pogarsky, G., Kim, K., & Paternoster, R. (2005). Perceptual change in the national youth survey: lessons for deterrence theory and offender decision-making. *Justice Quarterly*, *22*(1), 1–29. https://doi.org/10.1080/07418820420003336 27

Restivo, E., & Lanier, M. M. (2012). Measuring the contextual effects and mitigating factors of labeling theory. *Justice Quarterly*, *32*(1), 116–141. https://doi.org/10.1080/07418825.2012.756115

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Schulte, P. (2019). *Kontrolle und delinquenz: panelanalysen zu justizieller stigmatisierung und abschreckung*. Waxmann Verlag.

Schulz, S. (2014). Individual differences in the deterrence process: which individuals learn (most) from their offending experiences? *Journal of Quantitative Criminology*, *30*(2), 215–236. https://doi.org/10.1007/s10940-013-9201-6

Schwalbe, C. S., Gearing, R. E., MacKenzie, M. J., Brewer, K. B., & Ibrahim, R. (2012). A meta-analysis of experimental studies of diversion programs for juvenile offenders. *Clinical Psychology Review*, *32*(1), 26–33. https://doi.org/10.1016/j.cpr.2011.10.002

Sherman, L. W. (1993). Defiance, deterrence, and irrelevance: A theory of the criminal sanction. *Journal of Research in Crime and Delinquency*, *30*(4), 445–473. https://doi.org/10.1177/0022427893030004006

Slocum, L. A., Wiley, S. A., & Esbensen, F.-A. (2016). The importance of being satisfied: A longitudinal exploration of police contact, procedural injustice, and subsequent delinquency. *Criminal Justice and Behavior*, *43*(1), 7–26. https://doi.org/10.1177/0093854815609069

Snyder, H., & Sickmund, M. (1999). *Juvenile offenders and victims: 1999 national report*. U.S. Department of Justice.

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, *15*(3), 250–267. https://doi.org/10.1037/a0018719

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. https://doi.org/10.1214/09-STS313

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). CRC Press.

Ward, J. T., Krohn, M. D., & Gibson, C. L. (2014). The effects of police contact on trajectories of violence: A group-based, propensity score matching analysis. *Journal of Interpersonal Violence*, *29*(3), 440–475. https://doi.org/10.1177/0886260513505219

Wikström, P.-O. H. (2008). Deterrence and deterrence experiences: preventing crime through the threat of punishment. In S. G. Shoham, O. Beck, & M. Kett (Eds.), *International handbook of penology and criminal justice* (pp. 345–378). CRC Press.

Wikström, P.-O. H. (2011). Does everything matter?: addressing the problem of causation and explanation in the study of crime. In J. M. McGloin, C. J. Sullivan, & L. W. Kennedy (Eds.), *When crime appears: The role of emergence* (pp. 53–72). Routledge.

Wikström, P.-O. H. (2020). Explaining crime and criminal careers: The DEA model of situational action theory. *Journal of Developmental and Life-Course Criminology*, *6*(2), 188–203. https://doi.org/10.1007/s40865-019-00116-5

Wikström, P.-O. H., Mann, R. P., & Hardie, B. (2018). Young people's differential vulnerability to criminogenic exposure: bridging the gap between people- and place-oriented approaches in the study of crime causation. *European Journal of Criminology*, *15*(1), 10–31. https://doi.org/10.1177/1477370817732477

Wikström, P.-O. H., Oberwittler, D., Treiber, K., & Hardie, B. (2012). *Breaking rules: The social and situational dynamics of young people's urban crime*. Oxford University Press.

Wikström, P.-O. H., & Treiber, K. (2018). The dynamics of change: criminogenic interactions and life course patterns in crime. In D. P. Farrington, L. Kazemian, & A. R. Piquero (Eds.), *The Oxford handbook of developmental and life-course criminology* (pp. 271–294). Oxford University Press.

Wikström, P.-O. H., & Treiber, K. (2016b). Situational theory: The importance of interactions and action mechanisms in the explanation of crime. In A. R. Piquero (Ed.), *The handbook of criminological theory* (pp. 415–444). Wiley.

Wikström, P.-O. H., & Treiber, K. (2016a). Beyond risk factors: An analytical approach to crime prevention. In B. Teasdale, & M. Bradley (Eds.), *Preventing crime and violence* (pp. 73–87). Springer.

Wikström, P.-O. H., Tseloni, A., & Karlis, D. (2011). Do people comply with the law because they fear getting caught? *European Journal of Criminology*, *8*(5), 401–420. https://doi.org/10.1177/1477370811416415

Wiley, S. A. (2015). Arrested development: does the grade level at which juveniles experience arrest matter? *Journal of Developmental and Life-Course Criminology*, *1*(4), 411–433. https://doi.org/10.1007/s40865-015-0019-9

Wiley, S. A., Carson, D. C., & Esbensen, F.-A. (2017). Arrest and the amplification of deviance: does gang membership moderate the relationship? *Justice Quarterly*, *34*(5), 788–817. https://doi.org/10.1080/07418825.2016.1226936

Wiley, S. A., & Esbensen, F.-A. (2016). The effect of police contact: does official intervention result in deviance amplification? *Crime & Delinquency*, *62*(3), 283–307. https://doi.org/10.1177/0011128713492496

Wiley, S. A., Slocum, L. A., & Esbensen, F.-A. (2013). The unintented consequences of being stopped or arrested: An exploration of the labeling mechanisms through which police contact leads to subsequent delinquency. *Criminology; An Interdisciplinary Journal*, *51*(4), 927–966. https://doi.org/10.1111/1745-9125.12024

Young, C., & Holsteen, K. (2016). Model uncertainty and robustness. *Sociological Methods & Research*, *46*(1), 3–40. https://doi.org/10.1177/0049124115610347

Zhang, L. (1994). Peers' rejection as a possible consequence of official reaction to delinquency in Chinese society. *Criminal Justice and Behavior*, *21*(4), 387–402. https://doi.org/10.1177/0093854894021004001

Zhang, L., & Messner, S. F. (1994). The severity of official punishment for delinquency and change in interpersonal relations in Chinese society. *Journal of Research in Crime and Delinquency*, *31*(4), 416–433. https://doi.org/10.1177/0022427894031004004

Author Biography

Florian Kaiser is a doctoral candidate at the Institute of Criminal Law and Criminology at the University Muenster. His primary research interests include sanctioning research, explaining delinquent behavior, the study of social indicators, and quantitative data analysis.