The London School of Economics and Political Science

Procedural Justice Theory and the Black Box of Causality

Krisztián Pósch

A thesis submitted to the Department of Methodology
of the London School of Economics and Political Science,
for the degree of Doctor of Philosophy, London, August 2018

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 87,310 words.

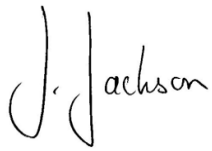
I confirm that Paper 3 was jointly co-authored with Jonathan Jackson, Ben Bradford, and Sarah MacQueen, and I contributed 45% of this work.

Signed:

Krisztián Pósch

As the candidates primary supervisor I hereby confirm that the extent of the candidate's contribution to the joint-authored paper was as indicated in the preface below.

Signed:



Dr. Jonathan Jackson

Acknowledgements

First and foremost, I would like to thank both of my supervisors, Jon Jackson and Jouni Kuha, as this thesis could not have been written without their complementary skills and mentorship styles. No one could ask for a better pair of supervisors, who helped me to think creatively about theory and methods whilst maintaining empirical rigour. A special thank you goes to Jon for being not only a brilliant adviser but also an academic role-model, who guided my transition from social psychologist to quantitative criminologist. His clear vision of the theory and his intellectual honesty aided me throughout the years in too many ways to describe.

I would like to express my gratitude to the faculty at the Department of Methodology, especially Dominik Hangartner, David Hendry, Benjamin Lauderdale, and Ben Wilson, who provided me with their feedback regarding my research, teaching, and professional development. In addition to my supervisors, their support was invaluable during the PhD. I would also like to thank Catherine J. Calogero, Patrik Korda, and Andrew Mitchell who not only helped with proofreading most of the materials in this thesis but also assisted in rewiring my brain in how to express my thoughts in academic English.

I also owe gratitude to the inspiring people in the LSE PhD community. I would like to thank Katharina Hecht, Christian Mueller, Thiago Oliviera, Tom Paskhalis, and Kohei Watanebe not only for inspiring me to strive for the best possible research, but also for providing a friendly working and teaching atmosphere. Thanks to you, this environment made it very easy to prosper both professionally and emotionally. Special thanks go to Katharina and Thiago for providing invaluable feedback for earlier versions of this thesis.

I would also like to thank the amazing friends I made in London through my passion for playing board games, especially Catherine J. Calogero, William-James Grey, Nicholas Kirkman, and Chris Warr. Saving the world from deadly diseases, vying for galactic domination, raising crops and pigs, or trading in the Mediterranean provided a welcome distraction and alternative mental puzzles to solve every week. They all made me feel welcome in an enormous and unwieldy city, and helped me to discover new places, people, and activities whilst providing more support than they can possibly imagine.

There are also people in my home country, Hungary, to whom I am equally grateful. I would like to express my gratitude to all of my friends, too numerous to mention here, for all the love and support they provided, despite being thousands of kilometres away most of the time. I also thank my parents, Andrea Kozáry and Gábor Pósch for all the encouragement and unwavering support they have given me over the years.

Finally, I would like to thank my love and soon-to-be wife, Júlia Varga for her love, her unconditional support, and for putting up with me and my workaholism throughout the years. Her advice and clarity of mind constantly helped me to find the right direction for my research, which makes all the achievements of this thesis as much hers as mine.

Krisztián Pósch, August 2018

Abstract

This thesis makes a theoretical and a methodological contribution. Theoretically, it tests certain predictions of procedural justice policing, which posits that neutral, fair, and respectful treatment by the police is the cornerstone of fruitful police-public relations, in that procedural justice leads to increased police legitimacy, and that legitimacy engenders societally desirable outcomes, such as citizens' willingness to cooperate with the police and compliance with the law. Methodologically, it identifies and assesses causal mechanisms using a family of methods developed mostly in the field of epidemiology: causal mediation analysis. The theoretical and methodological aspects of this thesis converge in the investigation of (1) the extent to which procedural justice mediates the impact of contact with the police on police legitimacy and psychological processes (Paper 1), (2) the mediating role of police legitimacy on willingness to cooperate with the police and compliance with the law (Paper 3, Paper 4), and (3) the psychological drivers that channel the impact of procedural justice on police and legal legitimacy (Paper 2). This thesis makes use of a randomised controlled trial (Scottish Community Engagement Trial), four randomised experiments, and one experiment with parallel (encouragement) design on crowdsourced samples from the US and the UK (recruited through Amazon Turk and Prolific Academic). The causal evidence attests to the centrality of procedural justice, which mediates the impact of an encounter with the police on police legitimacy, and influences psychological processes and police legitimacy. Personal sense of power, not social identity, is the causal mediator of the effect of procedural justice on police and legal legitimacy. Finally, different aspects of legitimacy transmit the influence of procedural justice on distinct outcomes, with duty to obey affecting legal compliance and normative alignment affecting willingness to cooperate. In sum, most of the causal evidence is congruent with the theory of procedural justice.

Submission plan

The four papers in this thesis have been or are planned to be sent to various journals.

An earlier version of Paper 1, titled *"Prying Open the Black Box of Causality: A Causal Mediation Analysis Test of Procedural Justice Policing"*, has been published as an LSE Legal Studies working paper in January 2018. This thesis contains an updated and reworked version based on the feedback received on that working paper, which has been submitted to the Journal of Quantitative Criminology.

Paper 2, titled *"It's nice to be empowered": An experimental assessment of psychological drivers of police legitimacy"*, is to be submitted to Criminology in the coming weeks.

The co-authored Paper 3, *"Truly Free Consent"? Clarifying the Nature of Police Legitimacy Using Causal Mediation Analysis"*, is planned to be submitted to the Journal of Quantitative Criminology.

Finally, Paper 4, *"Testing Complex Social Theories with Causal Mediation Analysis and G-Computation: Towards a Better Way to Do Causal Structural Equation Modelling"*, was submitted to Sociological Methods and Research, it has been through the RnR stage and is currently awaiting final decision.

Table of contents

List of figures

Introduction, conceptual overview, overview of the papers

Paper 1

Paper 2

Paper 3

Paper 4

Concluding remarks

List of tables

Paper 3

Paper 4

Preface

This document constitutes my PhD submission to the London School of Economics and Political Science. This thesis differs from the usual format because the Department of Methodology permits the "paper-based" model in its PhD Programme. Thus, the structure of the submission is somewhat unusual, and unavoidably entails some repetition of the theory, concepts, and methods discussed in each paper. The thesis starts with an introductory chapter which sets out a comprehensive model of procedural justice policing and police legitimacy and briefly overviews the importance of causal mechanisms in social explanation. The conceptual review introduces the key ideas and theories. It is followed by a brief empirical overview of each paper and the methods used within them. The four papers constitute the core of the submission but they are linked by short interludes to strengthen the narrative continuity of the thesis. Of these four papers, three are completely my own, but one (Paper 3) is co-authored with Jonathan Jackson, Ben Bradford, and Sarah MacQueen for which I contributed 45% of the work. Each paper is referenced separately as they would appear in a journal. The final chapter summarises what has preceded, outlines the main findings, discusses some limitations, and identifies future directions of research.

A PhD in Social Research Methods and Statistics is always a quirky endeavour. A thesis in methodology needs to be theory-driven, but it must also answer the arising substantive research questions with innovative tools and techniques. This chimaera-like nature means that this thesis comprises of two distinct components. First, relying on the existing criminological and psychological literature, it systematically tests the theory of police procedural justice and legitimacy. This is the organising force of the thesis, as each paper is motivated by questions and debates from the policing literature. Second, it uses causal mediation analysis, a family of methods mostly developed for epidemiology and biostatistics, to test and estimate the causal mechanisms predicted by the theory. Details of the various approaches to the statistical and design-based estimation of causal mediation analysis are mostly discussed in the papers, with Paper 1 and Paper 4 written as comprehensive reviews of causal mediation analysis with single and multiple mediators respectively.

Introduction

Why do people cooperate with the police? Why do people comply with the law? What should the police do to boost cooperation and compliance? These central questions of modern policing research need to be answered in order to develop effective policies of order maintenance, crime reduction, and civic engagement in communities. They are also inherently causal questions.

After a long period of pursuing such goals by focussing on instrumental motivators through toughening criminal sentences, using proactive and invasive policing techniques, broadening the purview of the police, and increasing police numbers, the last couple of decades have seen a marked shift in policy debate from coercive strategies to more consensual ones (Tyler, Goff, and MacCoun 2015). This shift was triggered by the perspective of procedural justice originally developed in social psychology by Tom Tyler and his colleagues (Lind and Tyler 1988; Sunshine and Tyler 2003; Tyler 2006b, 2011).

In a nutshell, the theory of procedural justice emphasises the importance of *how* people are treated instead of what kind of *outcome* they receive. When people perceive the decisions made by the authorities as neutral and fair, and when they are treated with dignity and respect, people in turn ascribe trustworthy motives to the power holders. Procedural justice activates value-driven self-regulation and makes people cooperate with authorities and comply with rules not out of concern for being punished or individual risk-benefit analysis but because it is *the right thing to do*.

It has also been argued that procedural justice does not directly predict cooperation and compliance but that it does so through citizens' evaluation of the legitimacy of the authorities (Tyler 2006a; Tyler and Jackson 2013). When people are treated in a procedurally just manner, they tend to find the authorities morally appropriate and give consent to their actions and demands even when they disagree with them. Hence, legitimacy is a property of authorities which makes people more likely to engage in societally desirable outcomes such as legal compliance or cooperation.

Yet, despite plentiful empirical support and the enormous influence of these ideas on modern policing research, there is only limited causal evidence which supports the theory and almost no assessment of the causal mechanisms hypothesised.

As observed by Nagin and Telep (2017 : 18) in a review of the literature to date: *"What has not been established is whether these associations reflect a causal connection between procedurally just treatment and perceived legitimacy and compliance."*

This limited evidence base should be concerning to researchers and policymakers alike. Without empirical research demonstrating robust causal relationships, it is difficult to devise successful initiatives. Moreover, the lack of knowledge regarding the causal mechanisms means that even if a causal relationship is established, the researcher cannot be certain *how and why* it arose. The prime aim of this thesis is to address this gap in the literature in two ways: (1) by introducing methods that are, when used together with appropriate data and research design, capable of testing causal mechanisms and (2) by using those methods to test the main hypotheses of the procedural justice literature.

To set the scene for the conceptual review and research questions, this introduction focusses on these two main aspects of the thesis. First, it proposes a new comprehensive model of procedural justice and police legitimacy. This comprehensive model is juxtaposed with other existing frameworks and the limitations and points of agreements between them are discussed. This framework is offered both as a theoretical and methodological tool for future theory development and hypothesis testing. Second, the introduction provides a brief overview of causal mechanisms in general and locates the place of causal mediation in particular. It is argued that there are disciplinary, theoretical, and statistical reasons that make this family of methods especially suitable for the present inquiry.

A comprehensive model of procedural justice policing

Several frameworks have been offered for procedural justice policing. One of the most elaborate frameworks provided by Hamm et al. (2017), is an apt springboard for the discussion of the current comprehensive model. As shown by Figure 1a, Hamm et al.'s framework starts with evaluations of the interactions including both proximate (e.g., procedural justice) and distal (e.g., effectiveness) assessments. These assessments shape the evaluation of the target (i.e., the police) and influence attitudes regarding their trustworthiness and normative alignment. In turn, these attitudes influence the internalisation of trust (in the police) and free deference to the authority (i.e., duty to obey). Finally, these internalised factors lead to reactions, such as willingness to cooperate with the police and compliance with the law.

While the proposed model of Hamm et al. (2017) contributes much to theory-development and empirical testing, I have four comments and recommendations. First, Figure 1a is a flowchart. It is not clear whether the effect of one construct on the next is exclusive, or it might have direct or subsequent downstream effects on the other elements of the theory. Including arrows to indicate the possibility of such effects is crucial for conceptual and modelling clarity, and also essential from a causal inference point of view, where the presence or absence of arrows represents hypothesised relationships or the lack thereof (Morgan and Winship 2014). With this in mind, Figure 1b offers an alternative model that incorporates such arrows in a way that all previous constructs are assumed to have an impact on all subsequent ones.

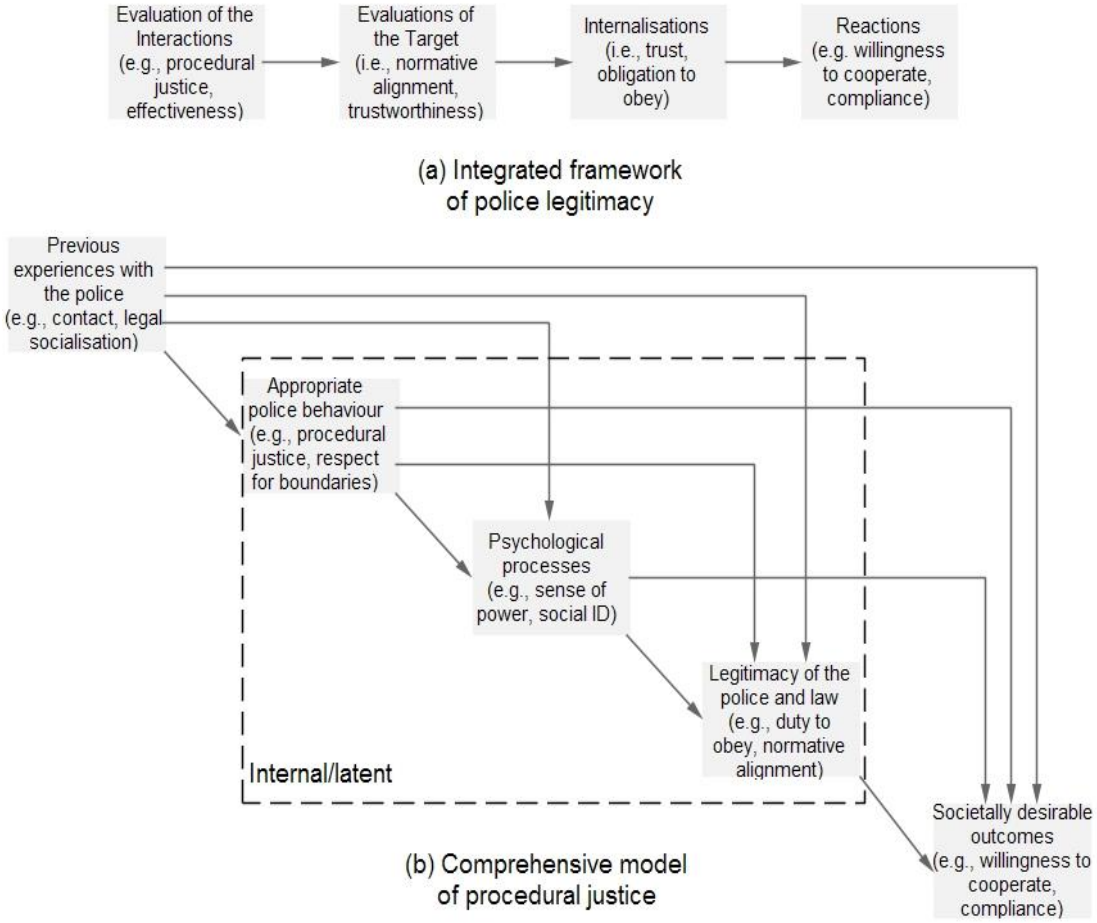


Figure 1 (a) Hamm et al.'s (2017) integrated framework and (b) the comprehensive model of procedural justice

Second, it is important to differentiate between the evaluation of the interactions with the police and the actual interactions, between the potentially

observable contexts, contacts, and behaviours during an encounter and the ways in which people judge that encounter. This distinction is germane as we know very little how encounters that are judged as procedurally fair by some are differentially evaluated by others (Nagin and Telep 2017). In Figure 1b "Previous experiences with the police" encompasses all the potential direct and indirect factors that might influence later components in the model. To provide a non-exhaustive list, such previous experiences include elements such as legal socialisation (e.g., Cavanagh and Cauffman 2017; Sindall, McCarthy, and Brunton-Smith 2016; Trinkner and Tyler 2016), contact with the police (e.g., Bradford 2017; Gau 2013; Myhill and Bradford 2012; Tyler, Fagan, and Geller 2014), vicarious contact with the police (e.g., Augustyn 2016; Gau and Brunson 2010; Rosenbaum 2005; Tankebe 2010), media effects (e.g., Desmond, Papachristos, and Kirk 2016; Gauthier et al. 2018; Hohl, Bradford, and Stanko 2010), effects of structural disadvantage (e.g., Bradford, Stanko, and Jackson 2012; Intravia et al. 2016; Kirk and Papachristos 2011), effects of immigration status (e.g., Bradford 2014; Bradford et al. 2015; Bradford, Jackson, and Hough 2016; Murphy and Cherney 2012), effects of cross-national differences (e.g., Bradford et al. 2014; Jackson et al. 2014; Johnson, Maguire, and Kuhns 2014; Tankebe 2009), and so on. As it is difficult to disentangle the effects of these elements (i.e., create a causal hierarchy), these are all listed under "Previous experiences with the police".

The next construct is "Appropriate police behaviour", which incorporates expectations regarding the procedural justice and distributive justice of the police, respect for boundaries, police effectiveness, and so on. It is followed by psychological processes, such as social identification, sense of power, emotions, and so on. All prior constructs are posited to influence the legitimacy of the police and the law, which can be conceptualised as normative alignment, duty to obey, legal cynicism, etc. Finally, the last construct in the model entails "Societally desirable outcomes", which include willingness to cooperate with the police, compliance with the law, community engagement, support for legitimate use of force by the police, and so on. These are all elements that will be further discussed in the conceptual overview.

To add to this second point, it is also important to highlight that in the comprehensive model only "Previous experiences with the police" and "Societally desirable outcomes" are potentially directly observable or manifest, the rest of the attitudinal processes are internal or latent. To highlight this distinction, the

unobservable processes that happen inside people's minds are in a separate box with a dashed line in Figure 1b.

My third comment concerns the motivation behind the model building pursued by Hamm et al. (2017). Their model synthesises the procedural justice and trust literature, showing parallels and equivalences between the two fields. This is a fruitful undertaking, as it encourages researchers from two distinct fields to engage with each others' concepts and work. However, merging these two theories is purely conceptual and lacks an overarching organising principle that could advise researchers how to integrate other elements into this model in the future.

One such organising principle proposed here is motivated social cognition (Von Hippel, Lakin, and Shakarchi 2005; Jost et al. 2003; Kruglanski 1996). This social psychological perspective proposes as central the idea that an understanding of human information processing is paramount when building attitudinal and behavioural models. Based on cognitive psychological research, it argues that people actively monitor (process) information, and that they have certain preferences which motivate them to initiate or withdraw from actively engaging with certain cognitive evaluations. This is bolstered by a mental architecture where certain psychological processes are pre-disposed to be quick and inaccurate (i.e., heuristics), whilst others require more time and deeper analyses (Kahneman 2012). For instance, when it comes to evaluation of the fairness or unfairness of a situation (Tabibnia, Satpute, and Lieberman 2008), or rule-following or rule-breaking (van Lier, Revlin, and de Neys 2013) human information processing tends to be very speedy, which implies that these are more basic psychological processes. In the procedural justice literature, this is usually referred to as a fairness heuristic (Lind, 2001; Proudfoot and Lind, 2015), which permits effortless and automatic (but often imprecise) processing of information. This heuristic can be challenged by an experience (a "teachable moment"), prompting controlled, systematic processing potentially updating the automatic processes (Tyler et al. 2014). In Figure 1b, these heuristic confirming/altering-events are listed under "Previous experiences with the police" and the heuristics are contained in "Appropriate police behaviour". Compared to the evaluation of fairness and legality, there are higher level psychological processes which transmit the effects of appropriate police behaviour on the evaluation of the legitimacy of the police and the law. In turn, views about the legitimacy of the power-holders shape observable behaviour. This motivated social cognition approach to procedural justice is instructive regarding the

various concepts' place within the model and has been recently advocated by other authors as well (Barclay, Bashshur, and Fortin, 2017; Jackson, Bradford, Brunton-Smith, and Gray, 2018).

Returning to Hamm et al.'s model (Figure 1a), although earlier theoretical work might suggest that "Evaluation of the target" and "Internalisation" should be handled separately, such a distinction needs to be substantiated by some form of justification in the cognitive information processing of the constructs. Because there is no such evidence in the extant literature – of which I am aware – I merged these two elements under "Legitimacy of the police and the law" instead of keeping them separate.

Finally, Hamm et al. (2017) only briefly allude to how their model could be reconciled with alternative frameworks. Comparing a new theoretical model to existing ones is important, however to demonstrate the flexibility of the comprehensive model, I juxtapose it with two popular alternatives (Figure 2). Tankebe and his colleagues (e.g., Tankebe 2013; Tankebe et al. 2016) define legitimacy as procedural justice, distributive justice, lawfulness, and effectiveness (Figure 2c). They argue that legitimacy (thus operationalised) predicts duty to obey or consent to the actions of the authorities. It is notable that, other than the difference in conceptualisation (i.e., what is legitimacy?), the comprehensive model can be easily integrated with this line of work. The constructs referred to as legitimacy are part of "Appropriate police behaviour" whilst duty to obey is one component of "Legitimacy of the police and the law".

Integration with Jackson et al.'s conceptualisation (e.g., Jackson et al. 2012; Tyler and Jackson 2013, 2014) is even easier, as most of the constructs discussed in the comprehensive model were informed by their previous studies, and thus they are equivalent (the particular model shown in Figure 2d was taken from Jackson (2018)). Hence, "Encounters with legal officials" and "Normatively grounded compliance and cooperation" are both subsets of "Previous experiences with the police" and "Societally desirable outcomes" from Figure 1b, whilst the other two constructs in the middle of the model only differ in labelling from "Appropriate police behaviour" and "Legitimacy of the police and the law". Other differences between the comprehensive model and the one depicted in Figure 2d are the absence of psychological processes in the middle and the arrow pointing from encounters to the normatively grounded outcomes.

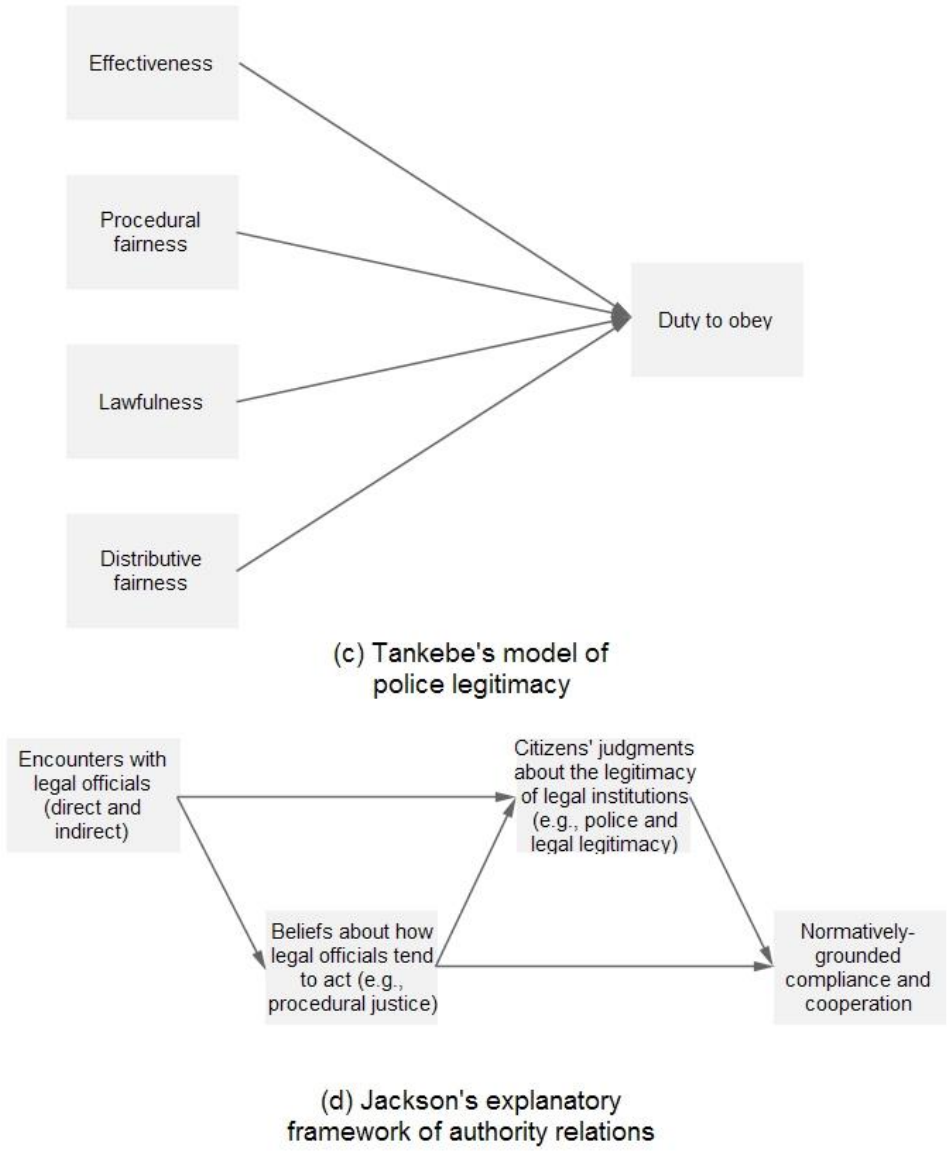


Figure 2 (c) Tankebe's (2012) model of police legitimacy and (d) Jackson's (2018) explanatory framework of authority relations

In conclusion, this new comprehensive model has at least four advantages: (1) it specifies all potential pathways, which is essential from a causal inference point of view and clarifies the hypotheses that could be tested, (2) it distinguishes between evaluations of interactions and actual interactions and latent and manifest realms for conceptual clarity, (3) it builds on a single guiding principle, motivated social cognition, which makes future model building possible, and finally (4) it can be easily reconciled with all three frameworks discussed in this introduction (Hamm et al. 2017; Jackson et al. 2018; Tankebe 2013).

Causal mechanisms

The second aim of this thesis is to empirically examine the pathways of the comprehensive model and to test the presence (i.e., causal or non-causal) and strength (i.e., sensitivity) of these effects, using a combination of appropriate methods and data. Because the comprehensive model involves multiple effects (i.e., directed arrows) it prompts the need to consider causal mechanisms. The importance of identifying causal mechanisms has come to the forefront of criminology in recent years. Sampson, Winship, and Knight (2013) selected causal mechanisms and pathways as one of the major challenges of translating research findings to policy. They acknowledged that randomised controlled trials (RCTs) possess strong internal validity and are usually successful in establishing a causal link between a particular treatment and an outcome. However, they also recognised that RCTs do not inform policymakers of *why* or *how* such interventions work. Yet, without specific knowledge regarding the underlying mechanisms, it is difficult to assess the viability of transferring policies from one context to another. In particular, Sampson, Winship, and Knight (2013) gave three reasons why causal processes are germane for policy analysis and implementation: (1) causal mechanisms can help differentiate between causes and confounders, (2) policy makers are generally concerned not only with simple causal effects but also with which route is taken during an implementation, and (3) policy efficacy usually demands consideration of alternative pathways of bringing about the same effect.

Matsueda (2017) also argued for analytical criminology to embrace the study of causal mechanisms. He stressed that methods can only become compatible with the goals of analytical criminology if they (1) allow for testing specific hypotheses and pathways without positing generative theories of causality, and (2) they make it possible to assess the social interaction effects and variables that are realised both on the micro (individual) and macro (aggregate, social) level alike. Specifically, he recommended the adoption of the potential outcome framework and the family of methods used throughout this thesis, causal mediation analysis, which can satisfy both of the aforementioned criteria.

Moreover, Kirk and Wakefield (2018) emphasised the importance of understanding and identifying causal mechanisms in their review of the collateral consequences of punishment. They encouraged future research to go beyond the correlational evidence and open the "black box of incarceration". They drew on a series of methods, from qualitative approaches (such as interviews and ethnographic

research) to social network analysis, to assess causal mechanisms. With such work, they hope to gain a better understanding of the effects of confinement on post-release outcomes and to identify certain intermediate mechanisms that can explain said outcomes.

Considering the newfound popularity of causal mechanisms in criminology, it is unfortunate that it is difficult to provide a single definition for causal mechanisms. This difficulty stems from the fact that such a definition needs to be rooted in one's understanding of the nature of causation itself. The complexity of such a task becomes apparent when browsing through the Oxford Handbook of Causation (Beebee, Hitchcock, and Menzies 2009), which dedicates ten chapters to the various standard and alternative approaches to causation. The multitude of accounts of causality is further obfuscated by the fact that competing explanations often complement each other even in a single discipline. For instance, most causal inference techniques applied in this thesis rely on counterfactual theories, but they also follow principles from probabilistic theories and causal modelling, whilst examining causal mechanisms is considered a separate philosophical approach in its own right.

This thesis does not directly engage with either of these theories other than by acknowledging the plurality of philosophical ideas in the field. This, however, means that I can only define causal mechanisms through the elements that are shared by these theories. Notably, Hedström and Ylikoski (2010) identified four characteristics that are shared by most explanations of causal mechanisms:

1. They are defined by the causal effect or phenomenon that produced them (e.g., evaluations of procedural justice are the product of previous experiences with the police).
2. A mechanism is a causal notion which refers to a process that produces the effect of interest (e.g., procedural justice produces an effect on police legitimacy).
3. All mechanisms provide structure, which makes the black box of causality transparent (e.g., previous experiences with the police affect attitudes towards police legitimacy through subjective procedural justice).
4. Causal mechanisms form a hierarchy, where certain effects by definition precede other effects. Although there might be a concern that such causal hierarchies could follow an infinite regress (e.g., previous experiences with the

police can be preceded by genetic or prenatal influences), in practice, it is reasonable to let the causal processes "bottom out" at disciplinary boundaries.

The main difficulty in defining causal mechanisms is that they have duel properties. They are descriptive elements of the causal process where they mediate the effect of a treatment on the outcome; this way they help to provide answers to *how* certain effects come about (e.g., "How does a procedurally just encounter increase police legitimacy?"). Simultaneously, they also produce the subsequent effect that helps to explain *why* certain treatments work (e.g., "Why does a procedurally just encounter increase police legitimacy?").

There are various different methods across disciplines that aim to tackle causal mechanisms. In criminology, studies have used qualitative methods to open the "black box" of causality by using interviews (Haberman 2016) and focus groups (MacQueen and Bradford 2017). There is also a growing literature outside of criminology which applies process tracing to identify causal mechanisms (Fairfield and Charman 2017; Saylor 2018). An increasing number of studies use mixed methods not only to establish causal pathways (Weller and Barnes 2016) but also to reconcile different concepts of causality and causal transmissions (Johnson, Russo, and Schoonenboom 2017).

Among quantitative methods, it is popular to harness relational information by using network analysis to confront causal mechanisms. In criminology, this perspective is exemplified by the works spearheaded by Papachristos and his colleagues (e.g., Papachristos et al. 2012; Papachristos, Wildeman, and Roberto 2015). Social network analysis allows the drawing of conclusions based on, for instance, group cohesion (density), an individual's place in a network (centrality), and an individual's levels of social interactions (peer influence, social contagion). Despite the illuminating findings of these studies, statistical work has been critical of drawing causal inference using similar methods of network analysis, finding them only appropriate in a very restrictive set of circumstances (VanderWeele, Ogburn, and Tchetgen Tchetgen 2012). Very recently, work has been carried out to take down such barriers, which makes this approach a promising alternative for the future (Ogburn et al. 2017).

With all these approaches considered, this thesis will use causal mediation analysis to tap into causal mechanisms (e.g., Imai et al. 2011; Keele 2015; Pearl 2001; VanderWeele 2015). For much of the work on causal mediation analysis, the potential

outcome framework is utilised as an explanatory tool to describe the causal effects. The idea of causal mediation analysis is very similar to the mediation analysis routinely used in Structural Equation Modelling (SEM) (Baron and Kenny 1986), as it postulates that if certain causal identifying assumptions are satisfied, the average treatment effect can be decomposed into direct and indirect effects. The indirect effect stands for the effect of the treatment that goes through an/multiple intermediate (mediating) variable(s) towards the outcome, whilst the direct effect stands for the remaining (unmediated) effect of the treatment on the outcome. In this approach, the indirect effect captures the causal mechanism as it describes how and why the treatment affects the outcome.

I concur with Matsueda (2017) that causal mediation analysis is capable of hypothesis testing without making law-like assumptions (i.e., that these mechanisms are bound to work) and that it can connect the individual and macro-levels in social explanations. Beyond these considerations, in this thesis I intend to demonstrate five more reasons why I think causal mediation analysis is well-equipped to test the comprehensive model of procedural justice outlined earlier:

1. This family of methods originates in SEM, which is a widely applied modelling strategy in the procedural justice literature, making the concepts of causal mediation analysis more familiar to criminologists working in this field.
2. The potential outcome framework provides much-needed clarity and rigour regarding the causal identifying assumptions.
3. Causal mediation analysis can be considered an analytical extension of estimating average treatment effects for RCTs or randomised experiments. Provided that the causal identifying assumptions are satisfied, it becomes feasible to decompose such average treatment effects into direct and indirect effects.
4. For several methods, sensitivity analysis techniques are readily available and can be used to quantify the robustness of the effects in relation to certain causal identifying assumptions.
5. Finally, these models improve upon SEM by providing more flexibility in modelling and looser causal identifying assumptions.

I end this introduction with a word of caution regarding causal mechanisms. Despite all the promise of studying causal mechanisms, I would advise against assessing mediating effects as an exploratory exercise. As in all rigorous scientific research, the first steps ought to be theory building, followed by empirical tests of associations, then average causal effects, and, only then should causal mechanisms be sought. Without extensive knowledge about the place of a construct in a broader theoretical model, it is difficult to tell whether emerging effects have true causal properties or can be explained by an influential unmeasured third source. Hence, causally testing mechanisms should only be carried out on mature theories that have been rigorously tested and bolstered by substantial empirical evidence. Fortunately, procedural justice policing can be considered one of these mature theories, where testing causal mechanisms is a natural next step.

Conceptual Overview

Why do people obey the law?

There are two major criminological accounts of why people obey the law and how legal authorities can encourage compliance. The classical, instrumental model considers people as rational-economic calculators who are mainly influenced by the certainty, severity, and celerity of punishments. Becker (1968, 1974) argued that the perceived losses associated with non-compliance can be increased through increasing (1) the likelihood of detention, (2) the severity of sanctions, and (3) the swiftness of justice. Accordingly, to encourage legal compliance, institutions need to focus on the potential deterrent effect of the criminal justice system by boosting the number of police officers, enacting stricter sanctions, and guaranteeing swift procedures. In other words, to manage crime they should pursue coercive "command and control" policing techniques (Hough 2012; Tyler et al. 2015; Tyler and Huo 2002).

Despite the worldwide popularity of this approach, there is mixed evidence on the success of the model. A meta-analysis on instrumental motivators found limited or negligible effects on crime control (Pratt et al. 2008). Another review of the empirical evidence (Nagin 2013) showed that, while police presence can have some deterrent effect, increasing the length of prison sentences seemed to give only a marginal effect and capital punishment appeared to have no effect at all. Other scholars (Charles and Durlauf 2013; Paternoster 2010) have highlighted some of the methodological and theoretical shortcomings of existing work, concluding that the evidence base is not strong enough to make a clinching judgement regarding the effectiveness of deterrence.

Doubts have also been raised about whether people can rationally assess the risks of punishment. Kleck and Barnes (2008) found no evidence for the claim that people have a realistic idea about the chances of being apprehended, either on the individual or the aggregate (collective) level. Indeed, research participants showed self-attribution bias, over-estimating their personal ability to avoid arrest. In another study, Holliday, King, and Heilburn (2013) found that while convicted felons had a realistic view about the likelihood of apprehension, their judgment regarding their own personal chances suffered from positivity bias resulting in the belief that they are "exceptions to the rule". Thus, even if people hold realistic views about the probability

of punishment, those views might not predict how they think about their own prospects, making it difficult for legal authorities to alter perceptions.

Despite such reservations, several studies have demonstrated (e.g., Sunshine and Tyler 2003; Tyler 2006b; Tyler and Jackson 2014) that perceived risk of sanction predicts self-reported offending behaviour, even though it has limited explanatory power. Thus, the debate has recently shifted away from questioning the influence of deterrence towards a critical account of the potentially harmful social and economic side-effects of policing that follows these principles (Tyler et al. 2015). Several studies have revealed increased costs to the criminal justice system due to the elevated number of police officers, the increased use of surveillance techniques, and the hiring of private security to monitor public spaces, such as subways or schools (e.g., Durlauf and Nagin 2011; Punch 2007; Wacquant 2009). Others have pointed to issues associated with the increased police presence in and wider scrutiny of disadvantaged neighbourhoods, which can increase the ghettoisation of such areas and the system avoidance of locals, thus maintaining their marginalised position in society (Brayne 2014; Goffman 2009) and evoking the Foucauldian concept of panopticism (Foucault 1977). As Kirk and Wakefield (2018) have shown, it is difficult to assess thoroughly the collateral consequences of criminal justice engagement, but they include among other things diminished physical and mental health, declining employment prospects, reduced civic engagement, housing and residential instability, and so on.

In this thesis, I build on an idea that is gaining currency in the academic and political debate: namely, that legitimacy may be a viable, consensus-based alternative to the coercion-based model of crime-control. According to this perspective, compliance with the law is not primarily informed by the potential risk of being caught but is instead predicted by public perceptions of the morality, fairness and authority of the agents of the justice system. By contrast to the rational choice model that is aligned with deterrence, legitimacy offers a value-based explanation, which posits that people obey laws principally because of their personal, internalised normative incentives (i.e., they think that obeying the law is the "right thing" to do, see: Tyler 2006a; Tyler and Jackson 2013). Legitimacy can surpass self-interest because it relies on relational mechanisms that underpin the social relationships between power-holders and subordinates (Jackson and Gau 2015; Tyler and Blader 2003; Tyler et al. 2015; Tyler and Lind 1992).

Why are the police important for legitimacy?

Considering the influential nature of the concept of legitimacy, it is not surprising that in the literature there are many different uses and forms of legitimacy, distinguished by their level (instrumental vs. personal) and their particular field (political, governmental, etc.) (Abulof 2016; Hinsch 2010). This thesis argues that there are compelling reasons to focus mainly on police legitimacy. The police are the most accessible and visible agents of the justice system, representing in some sense the people they are policing (Jackson and Bradford 2009). As noted by Tyler and Huo's (2002) classic book, which focussed on data from the United States, while 44% of the respondents reported that they had encountered a police officer in the previous two years, only 8% had attended a court. The police may thus play a prominent role in forming public attitudes regarding the rule of law. Indeed, police officers have been described as "street corner politicians", who take on an intermediary role between power-holders and subordinates in expressing prevailing changes in laws and regulations to the public (Muir 1977).

Contrary to the stereotypical crime-fighting image of policing, the police's major role is not to gather intelligence and solve crimes but to maintain social order and reassure people that when help is needed they can rely on a safety net. This role is described by Reiner (2010, 2012) as "fire brigade policing" or "first aid order maintenance". Police officers are continuously sending signals of authority and are capable of upholding, boosting, or eroding perceptions regarding the quality of governance (Loader 2014), with a particularly strong impact on views of the legal system as a whole (Baker et al. 2013). Police officers are "condensation symbols" for the state and, through their activities, they help people to make sense of and give order to the social world (Loader 2006). In line with this representative function, police officers not only express but also form and maintain the social status of a society, in other words, they "patrol the boundaries of social identities [of the individuals']" (Bradford 2014: 24).

Procedurally just policing and contact with the police

Since Tyler's (2006b) ground-breaking work, and its later extension to the policing context (Sunshine and Tyler 2003), it has become widely accepted that the foremost constituent of popular views regarding appropriate police behaviour in Western democracies and the primary factor underlying police and legal legitimacy is whether

people believe that the police are procedurally fair. While various studies have conceptualised procedural justice differently, a good deal of consistency has emerged. Tyler and Jackson (2013) argued that procedural justice entailed three elements of (1) voice, (2) neutrality, openness, or transparency, and (3) dignity, politeness, or respect. Mazerolle et al. (2013a, also see: Higginson and Mazerolle 2014; Mazerolle et al. 2013b) in a systematic review identified the four components of procedural justice as (1) citizen participation, (2) neutrality, (3) dignity and respect, and (4) communicating trustworthy motives. Recent studies by Tyler and colleagues (Trinkner, Jackson, and Tyler 2017; Tyler et al. 2014) grouped the aforementioned features into two categories of fairness of treatment (incorporating dignity, politeness, respect, and trustworthy motives) and fairness of decision making (incorporating giving voice, neutrality, openness, and transparency).

One of the key questions in the literature is how views regarding police procedural justice are shaped. The introduction of this thesis listed several elements that belong to previous influential experiences with the police and some of which can predict evaluations of procedural justice. Yet, for the police, the most important factor is how police practice itself can change the views of the police during encounters with the public. Understanding the effects of police-citizen encounters has become increasingly important, particularly in the United States, which has seen a shift in recent decades from reactive to proactive policing tactics, meaning that there is an increasing likelihood that the police-citizen encounters are not citizen- but police-initiated (Tyler, Jackson, and Mentovich 2015). Such encounters involve direct experiences (e.g., police contact) and indirect experiences (e.g., vicarious experiences, mass media).

Starting with direct contact with the police, there is a good deal of evidence that people who have recently had direct experience with the police have more negative views about the police and the justice system in general (Bradford, Jackson, and Stanko 2009; Skogan 2006). Reiner (2010) has argued that such findings can be partly explained by the fact that people are likely to meet officers during low points of their lives or when they have just experienced something horrific, been victimised, or were in an otherwise vulnerable condition because of which they required the help of the police. It follows that the police tend to be most trusted by people who do not usually meet them either because they do not require their services or because they live in areas where the police are less needed (Bradford 2017; Bradford et al. 2012).

Importantly, however, when scrutinising the variation in views among those who have had contact with the police, there is some evidence that the police's perceived procedural justice has a positive association with police legitimacy and compliance with the police. For instance, McCluesky's (2003) research on police encounters in Florida and Indiana found that the less coercive the police acted during encounters, the more likely that the citizens complied with their requests. Procedurally just treatment is also capable of informing the reason for an encounter (e.g., "Why was I stopped?"), for instance, by reducing the likelihood of evaluating a police stop as motivated by racial profiling (Tyler and Wakslak 2004). Moreover, even when police officers deliver negative outcomes, such as administering a fine, fair treatment can still be related to greater legitimacy (Tyler and Fagan 2008). Likewise, studies from Australia (Mazerolle et al. 2013a), Turkey (Sahin et al. 2017), and from the United States (Tyler et al. 2014) have found that procedurally just policing is linked to higher confidence in the police either with regards to previous encounters or in general. However, other studies have found contradictory evidence. Road-side stops of motorbikes (Gau 2013), cars (Epp, Maynard-Moody, and Haider-Markel 2014; Gau and Brunson 2012), and stop and frisk policies (Gau and Brunson 2010) all can have adverse effects on police legitimacy; in fact, they sometimes even increase the likelihood of future delinquent behaviour (Wiley, Slocum, and Esbensen 2013).

What explains this sizeable variation in the results of evaluating police encounters? Some recent scholarship has argued that police contacts should be considered in terms of patterns rather than discrete events. Tyler, Fagan, and Geller (2014) found that the positive effects of procedurally just policing tactics decreased following elevated exposure to police contacts, and that the police were judged more effective when they were perceived to engage in fewer street and car stops in a neighbourhood. Repeated police stops can accumulate in effect, amplifying distrust between officers and civilians, especially if some groups are disproportionally overrepresented among the stopped (Bradford 2017; Epp et al. 2014). Another study (Slocum, Ann Wiley, and Esbensen 2016) found that, while favourable police contact may be able to ameliorate negative effects, it cannot completely counter their association with future delinquent behaviour. Thus, although procedurally just policing may be able to minimise the negative effect of a contact, the dampening effect may not be total.

Another difficulty with police stops is that often there is an asymmetry in citizens' evaluations of the stops, where procedurally just contacts produce either no or minimal positive/negative correlations, whilst procedurally unjust contacts produce outsized negative associations with police legitimacy (Bradford 2017; Skogan 2006). This all indicates that the aim of procedurally just policing tactics in police stops is to limit the potential harm that a negative encounter might cause instead of increasing citizens' perception of the legitimacy of the police. Moreover, procedurally just practice is not the sole predictor of increased legitimacy; a right balance needs to be struck between the intensity and volume of police stops as well.

Finally, and as suggested previously, indirect contacts can also be important. Some scholars have argued that vicarious experiences can be more important than direct contact (Rosenbaum 2005), others have attributed a similar importance to them (Tankebe 2010), while others still have disputed the importance of vicarious encounters (Van De Walle 2009). Sources of indirect contacts include the media and advertising. Hohl, Bradford, and Stanko (2010) effectively used "leaflet encounters" to inform the public about police work and to convey procedurally just messages. In another study, Desmond et al. (2016) found that media coverage of violent police practices can have a negative impact on cooperation with the police. Overall, Mazerolle et al.'s (2013b) meta-analysis found that the particular vehicle of police intervention (direct or indirect) was less important, and the effect of an intervention was mainly influenced by whether it was perceived as procedurally just or not.

In this thesis, two papers (Paper 1 and Paper 3) rely on data from the Scottish Community Engagement Trial (MacQueen and Bradford 2015, 2017) where the behaviour of police officers in roadside police checks was manipulated. Instead of scrutinising the impact of the contact directly, both studies examined the mediated effects and aimed to explain the causal mechanisms that play a role in similar encounters.

In addition, and as an extension of the procedural justice framework, Paper 2 and Paper 4 measure and manipulate the subjective police respect for boundaries. Respect for boundaries is an expectation regarding appropriate police behaviour; it is the perception of whether police officers act within the boundaries of their rightful power and legitimate authority (Huq, Jackson, and Trinker 2017; Trinkner et al. 2017; Trinkner and Tyler 2016). For instance, some people might find repeated occurrences of stop-and-search intrusive, even if the police act in a procedurally just way in each

occurrence (Tyler et al. 2015). This could breach legal boundaries as perceived by the individual, and affect police legitimacy above and beyond procedural justice. In Paper 2 and Paper 4, experimental conditions are used where the police are not only procedurally unjust, but breach the perceived legal boundaries by engaging in illegal police practices. The goal of this thesis is not to disentangle respect for boundaries from procedural justice, but to test how procedural justice and respect for boundaries change in tandem given a certain experimental manipulation.

Legitimacy of the police

Regardless of its conceptualisation, procedural justice tends to outstrip concerns regarding distributive justice or the effectiveness of the police in the context of police legitimacy in Western countries. There has been a good deal of discussion recently regarding the meaning and measurement of police legitimacy (among others: Bottoms and Tankebe 2012; Hough, Jackson, and Bradford 2013; Huq, Jackson, and Trinker 2017; Jackson et al. 2014; Jackson and Gau 2015; Tankebe 2013; Tyler and Jackson 2013). Since police legitimacy is at the heart of this thesis, I discuss the alternative theoretical viewpoints starting with Hough, Jackson and Bradford (2013, Bradford et al. 2012; Jackson et al. 2012, etc.).

Beetham's (1991, 2013) account of legitimacy sought not only to understand laypeople's perception of legitimacy and their ensuing behaviour but also to establish the normative justification of power. According to him, legitimacy is rooted in:

(1) *legality*, thus power is exercised following established rules,
(2) *normative justifiability*, thus the established rules are accepted by both the power holders and subordinates, and
(3) *expressed affirmation/recognition*, thus the subordinates give their consent and authorise the power holder.

Whenever any of these elements becomes threatened, it translates to a system which is either illegitimate (illegality), struggles with a legitimacy deficit (lack of shared values), or suffers from delegitimation (withdrawn consent). Hough, Jackson, and Bradford (2013) operationalised these concepts as "lawfulness/legality", "normative alignment", and "obligation to obey the law" and treated the latter two as constituents of legitimacy.

Normative alignment refers to the belief that police officers act in ways that accord with societal expectations regarding the appropriate use of power, such that the institution they represent is deserving of the power it holds (and represents values that citizens think are important). Normative alignment stems from one of the core psychological goals of making attributional claims regarding other people's intentions based on their innate, unobservable characteristics. Since early psychological studies, it has been established that if people perceive others to be alike, then they will be readier to engage and identify with them (Heider 1958). To believe that authorities are benevolent, reasonable, and acting for one's benefit is to believe that authority figures generally act in normatively appropriate ways. Notably, normative alignment has been discussed in other places somewhat differently. Braga et al. (2014) call the same phenomenon identification with the police, emphasising the felt connection with officers. Trinkner, Jackson, and Tyler (2017) argue that when people believe that police officers act in normatively appropriate ways, this strengthens the sense that police officers share salient values, generating a broader sense of the normative justifiability of power.

By contrast, the second element of police legitimacy, obligation to obey, originated in Weber's (1922/1998) works. He suggested that individuals usually do not succumb under constant harassment of the state, instead, they adopt a sense of responsibility or obligation that it is morally just to obey the authorities, even when their intuition would advise otherwise. Thus, obedience becomes a positive civic duty which enables the authorities to reach a "Geltende Ordnung" (order by normative authority). Importantly, in the police legitimacy literature, this obligation is posited to be internalised and ensured by voluntary consent and not coercion, and it is assumed that people realise the boundaries between consent and coercion, thus obligation to obey is not blind submission towards the authorities (Beetham 1991; Tyler and Jackson 2013). Crucially, duty to obey is content free because individuals authorise the power-holders to dictate appropriate behaviour through this internalised sense of voluntary deference (Jackson and Gau 2015).

It is important to acknowledge that – as recognised by many (Bottoms and Tankebe 2012; Hinsch 2010) – Weber did not immerse himself in the idea of normative legitimacy. Although he acknowledged that an authority can be normative, he took a relativistic stance on what those norms ought to be. Admittedly, he believed that even an authoritative, totalitarian regime can be experienced as legitimate (Weber

1922/1998). As Bottoms and Tankebe (2012) emphasised, this leaves the door open for a kind of "dull compulsion" that does not rely on normative justification. They speculate that people might obey the police for other reasons (e.g., fear of the police) and researchers have raised (though not confirmed) whether this might apply in low trust contexts, such as Ghana (Tankebe 2009) or Trinidad (Johnson et al. 2014). If people do obey the police for reasons rooted in self-interest or immoral personal agendas, then – using the terminology of Raz (2009) – the authority should be understood as de facto instead of legitimate.

By contrast, Beetham (2013) maintained that legitimacy has substantive elements to it, basic principles which are beyond disagreement. The ideas outlined by Beetham follow a contractualist approach similar to the propositions by John Rawls (1999), postulating that there are universal values (such as the universal declaration of human rights) that inform people's understanding of the moral duty to obey. Beetham also clarified that the sheer fact that certain people would not obey laws for normative reasons does not mean that those reasons are refuted. This only reflects the well-known fact that vile motives are also present in human nature (Zimbardo 2007).

The most striking aspect of this theoretical discussion is the lack of empirical evidence addressing this debate. If there is indeed a separate normative and non-normative understanding of duty to obey, then these two constructs ought to be measured and tested with regard to other aspects of legitimacy and procedural justice. One of the contributions of this thesis is that, in co-authored Paper 3, we address this gap in the literature.

In contrast to the conceptualisation and operationalisation discussed so far, Tankebe (Bottoms and Tankebe 2012; Tankebe 2013; Tankebe et al. 2016) has proposed an alternative model of police legitimacy. He posits that police legitimacy is comprised of shared moral values, specifically the perception of police effectiveness, procedural justice, legality, and distributive justice. He theorises that the major (non-normative) outcome of legitimacy is felt obligation to obey (consent) which, in turn, inspires societally desirable outcomes such as compliance, cooperation, and so on.

As noted in the introduction, Tankebe's model can be easily reconciled with the comprehensive model proposed by this thesis by drawing an equivalence between Tankebe's (2013) concept of police legitimacy and appropriate police behaviour. Why Tankebe's conceptualisation of police legitimacy is not acceptable from the perspective of the comprehensive model is explained by its guiding force: motivated

social cognition. In all the recently discussed philosophical literature (Beetham 1991; Raz 2009; Weber 1922/1998), these otherwise differing perspectives concur that the perception of legitimacy is something considered and evaluated that requires careful deliberation. In comparison, psychological studies imply that at least the perception of justice and legality are automatic processes rooted in heuristics (Bell and Buchner 2012; van Lier, Revlin, and de Neys 2013; Lind 2001; Proudfoot and Lind 2015) – although I am not aware of any such studies on attitudes towards effectiveness. Thus, the psychological evidence implies that calling this group of attitudes legitimacy is unlikely to hold true and that these elements are better referred to as components of appropriate police behaviour.

A further issue with Tankebe's alternative model of police legitimacy has to do with the measurement of the constructs. There are some contradictions within the scales; for instance, Tankebe (2013) measured lawfulness both as lack of rule violation ("When the police deal with people in my neighbourhood, they always behave according to the law") and normative alignment with the law (i.e., shared values, "The law represents the moral values of people like me"). Further complicating the operationalisation of legitimacy, in cross-national works, Tankebe has argued that it might be impossible to use the same measurement model due to the special circumstances in countries such as Ghana (Tankebe 2009, 2010). This is in line with Bottoms and Tankebe's (2012) concern that the normative preconditions are likely to vary across social, cultural, political, and other contexts. However, Tankebe, in his own comparative work, continued using the same pre-determined measurement structure outlined earlier (Tankebe et al. 2016). Thus, there is a clear contradiction between the two viewpoints which appear to argue simultaneously that police legitimacy is likely to be context-dependent but has the same measurement structure regardless of where it is measured. For the progress of the theory of procedural justice and police legitimacy, it is essential to find agreement on how the different constructs should be measured while also maintaining conceptual clarity within the scales and with regards to their applicability (Jackson 2018).

To conclude, I now clarify the theoretical position and relationship between police procedural justice and police legitimacy relying on the discussion of Jackson and Gau (2015). They have argued that trust in the police encompasses expectations about the current and future behaviour of the police. This trust serves as an approximation of the intentions and capabilities of officers. Trust always requires a

"leap of faith" from the individual, as one cannot be certain how a power-holder will act in any prospective encounter. Hence, to make this critical judgment, a person relies on normative expectations that will involve perceptions of fairness, neutrality and so on (procedural justice), anticipated performance (effectiveness), the police's respect for playing by established rules (legality/respect for boundaries), and other similar properties and characteristics. These components together are referred to as appropriate police behaviour in the comprehensive model.

By contrast, legitimacy describes the rightfulness of power through moral appropriateness and a positive but nevertheless content-free duty to obey the police. As mentioned earlier, legitimacy is relational in nature; it defines the power-relations between the authority and subordinate. Trust and legitimacy can conceivably overlap in that they both describe the appropriate code of conduct based on the established norms, which represents normative alignment.

Another important distinction between legitimacy and trust is their level of influence. Trust is always a judgment formed on a micro-level and it governs interactions between individuals and the officers. Conversely, legitimacy is an institutional property that establishes whether the police have a right to exercise their power over the public. Even though they are at different levels, they are interdependent and shaped by an ongoing dialogue between the police officer (power-holder) and civilian (subordinate).

Recently some scholars have posited that there is a hierarchy within police legitimacy, normative alignment informing willing duty to obey the police (Hamm et al. 2017; Huq et al. 2017). As a test of this proposition, Paper 4 juxtaposes two solutions. In the first, normative alignment and duty to obey mutually reinforce each other and mediate the impact of procedural justice on willingness to cooperate with the police. In the second, a sequential order is assumed with procedural justice affecting duty to obey through normative alignment, but not the other way around.

Procedural justice, police legitimacy, and societally desirable outcomes

As argued earlier, procedurally just policing can assist authorities reaching their respective goals through boosting police legitimacy. A variety of different studies have linked legitimacy to compliance with the law (e.g., Jackson et al. 2012; Murphy, Tyler, and Curtis 2009; Sunshine and Tyler 2003; Tyler and Huo 2002), cooperation with the police (e.g., Hamm et al. 2017; Mazerolle et al. 2013a; Moravcová 2016a; Murphy and

Cherney 2012; Tyler and Fagan 2006), greater community engagement (Tyler and Jackson 2014), and normatively bounded beliefs regarding the acceptable use of violence in society (Bradford, Milani, and Jackson 2017; Jackson et al. 2013). Police legitimacy's scope may even be broader, with some recent studies linking legitimacy to greater support for counterterrorism initiatives (Cherney and Murphy 2013; Huq, Tyler, and Schulhofer 2011; Madon, Murphy, and Cherney 2017; Murphy, Cherney, and Teston 2018). As such, legitimacy seems to play a role in law-abiding behaviour, better flow of intelligence, and increased willingness from citizens to empower legal authorities.

Recent reviews and empirical tests in the literature (Jackson 2018; Tyler and Jackson 2013, 2014) have found that the different aspects of legitimacy are more likely to be associated with certain outcomes. Accordingly, duty to obey has been found to be more strongly associated with deferential outcomes, such as legal compliance and the acceptance of justified use of force. By contrast, normative alignment has been more closely related to proactive outcomes, such as willingness to cooperate and community engagement.

In this thesis, Paper 3 examines the mediated effects of duty to obey on both legal compliance and cooperation, whilst Paper 4 considers both aspects of legitimacy as mediators of the impact of procedural justice on cooperation.

Procedural justice and police legitimacy when legitimacy is challenged

Another strength of the procedural justice framework is that it is also capable of explaining why people turn away from the traditional, state-guaranteed delivery of justice, support vigilantism, and allow the emergence of alternative authorities. Recent studies (Bradford, Milani, et al. 2017; Jackson et al. 2013) have found that legitimacy is positively associated with the support for the normatively justified use of force by the police and decreases the acceptance of other sorts of public violence. By contrast, legitimacy deficit or delegitimisation can embolden people to take justice into their own hands as a protest against illegitimate government practices, as can be observed in Bolivia (Goldstein 2003), Mexico (Zizumbo-Colunga 2017), or Jamaica (Reisig and Lloyd 2008). Nivette (2013) proposed a plausible theoretical model of how the effects of illegitimate state actions might spiral through society (Figure 3). She posits that negative state actions have a pernicious impact on both individuals' attitudes and actions, which in turn evolve into detrimental effects on the state level. Nivette (2016)

herself tested the proposed theory across eighteen Latin American countries and found that legitimacy decreases, and police corruption increases, support for extra-legal capital punishment (e.g., revenge killings), while institutional effectiveness and statelessness (measured through homicide rates) only predicted one particular type of vigilantism. Importantly, Western countries are not immune to similar effects: high legal cynicism and low legitimacy have also been associated with increased offending behaviour and vigilantism (e.g., Kirk and Papachristos 2011; Penner et al. 2014; Wilkinson, Beaty, and Lurry 2009).

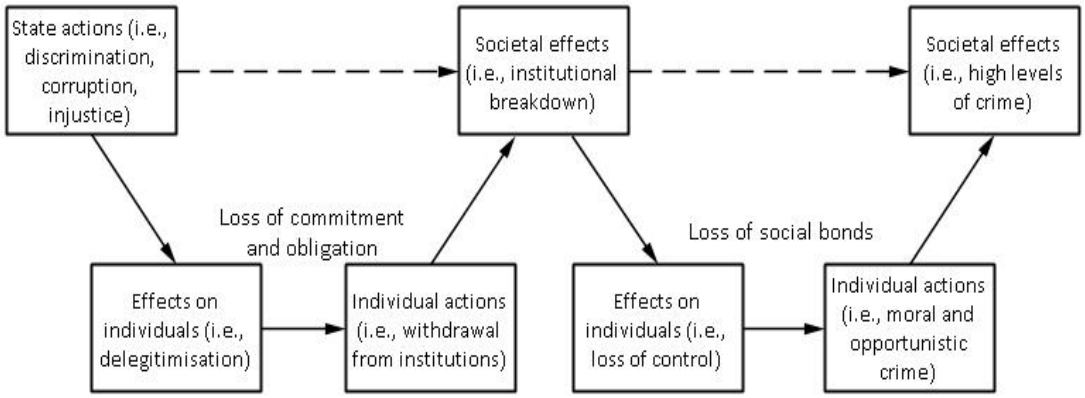


Figure 3 Nivette's (2013) theoretical model of the role of the state and its prediction of crime

The question still remains whether procedurally just practices are sufficient to restore legitimacy. Waddington et al. (2015) speculated that negative prior events might make it impossible for procedurally just practices to bear fruit. Their focus group study selected extreme cases of high distrust of the police, in which participants had seemingly unshakable negative convictions about officers and the institution. Tyler, Fagan, and Geller (2014) suggested the same and found that contextual information about a neighbourhood might improve or degrade perceptions of officer encounters. Augustyn's (2015) longitudinal study also suggests that previously held attitudes about procedural justice have long-lasting effects, and can even have an impact one or two years in the future. En masse, the studies underlining the importance of context verify Tyler's (2006b) initial presumption that trust is "motive-based" (i.e., the initial attribution of the police behaviour is decisive when judging an equivocal situation). At the same time, these findings do not tell us whether procedurally just interventions can

"break the ice", they only indicate that some people hold stronger views against the police.

Certain other recent studies have provided some succour in this regard. Bianchi et al. (2015) found that when trust was low among participants, then procedural justice (with the complementary but secondary effect of outcome fairness) had a much larger impact than when the trust was high. Correspondingly, their results also indicate that when trust is high, then a lack of procedural justice or outcome fairness will have more detrimental effects than when trust is low to begin with. In a similar vein, Kochel (2016) found that procedural justice remains a powerful predictor of legitimacy in areas with low collective efficacy and that aggressive policing (a mixture of police satisfaction and legality) becomes significant in such areas, but not in high collective efficacy ones.

This thesis mainly focusses on average direct and indirect effects, and the randomisation in all papers makes preconceived notions regarding procedural justice on average the same for each experimental condition. However, both Paper 2 and Paper 4 manipulate procedural justice and create artificial conditions in which the police are perceived as procedurally unjust or in breach of legal boundaries. Artificially created, these conditions are expected to serve as analogues to other contexts where procedural justice and legitimacy are challenged. The comparison of procedurally just and procedurally unjust/breach of boundaries conditions aims to approximate the effect of expecting fair or unfair/illegal treatment by the police.

Psychological processes connecting procedural justice to police legitimacy

When it comes to psychological processes regarding the perception of procedural justice and legality, it is worth reiterating and elaborating on what has been argued before. In line with cognitive psychological research and the motivated social cognition approach (Barclay et al. 2017; Von Hippel et al. 2005; Jackson et al. 2018; Jost et al. 2003; Kruglanski 1996), the human mind is equipped with heuristics which enable quick and effortless but usually imprecise ways of processing information on certain events, behaviours, and characteristics. By contrast, people are also capable of controlled, systematic processing, which is time-consuming, requires focus, but which also usually increases the accuracy of memories and future recall, and can inform and adjust the aforementioned heuristics. This line of research has established that the perception of procedural and distributive fairness (Lind 2001; Proudfoot and Lind

2015; Tabibnia et al. 2008) and legality and rule-breaking (Bell and Buchner 2012; Bonnefon, Hopfensitz, and De Neys 2013; van Lier et al. 2013b) are usually guided by heuristics. However, this literature has also shown that people are sensitive to information about violations of fairness or wrongdoing, which boosts awareness and prompts systematic processing.

Crucially, concerns regarding procedural justice become pertinent when power-differentials are salient, whilst in non-hierarchical settings procedural justice has limited or mixed effects (Mentovich 2012; van Prooijen, van den Bos, and Wilke 2002). Therefore, several studies have focussed on associations between procedural justice and personal sense of control, power, and power differentials all of which could potentially inform police legitimacy. In their foundational book, Thibaut and Walker (1975) scrutinised the social exchange between individuals and authorities. They argued that when people interact with power-holders they want to have a (perceived) influence over the outcomes of the interaction. Their work on the courts demonstrated that when individuals are allowed to voice their concerns and opinions they are more likely to accept the results of the proceedings and find them favourable, regardless of the actual outcome. In a similar vein, van Prooijen's (2009) work on autonomy-regulation and Mentovich's (2012) work on power and control concurred that procedural justice increases a sense of empowerment and sense of control over outcomes in contexts where power was unevenly distributed. This increased sense of control contributes to better behavioural inhibition (e.g., compliance) and a higher likelihood of behavioural engagement (e.g., cooperation) as well.

It has also been shown that procedural justice can reduce the subjective threat of uncertainty (Van den Bos 2001). When procedurally just principles are established, people find it easier to cope with uncertainty, as they perceive hierarchies governed by such standards as more stable and dependable (Hays and Goldstein 2015). It follows that the experience of empowerment is partly attributable to people being less bounded by authorities, as they have clear guidance on how to manage expectations of the future. Importantly, while power only means that an individual or institution has control over certain critical resources, procedurally fair power ascribes status to the power-holder (Blader and Chen 2012). This status engrains legitimacy, as the power-holder with status invokes (1) respect and admiration, (2) voluntary deference (i.e., compliance in the absence of threat or coercion), and (3) the assumption that the authorities provide instrumental social value (Anderson, Hildreth, and Howland 2015).

Adapting these findings to the policing context, by being neutral, showing respect and allowing voice the police establish procedurally just principles. These principles define the power-relations between the power-holder and the individual and increase the individual's subjective control, autonomy, and power in their future dealings with the police. In addition, procedural justice also signals stability, and reduces uncertainty and anxiety regarding future encounters, partly due to the perceived increase in mastery described earlier. Finally, through this empowerment, people are more likely to assign status and legitimacy to authorities, show respect, and engage in cooperation and voluntary deference because they believe that doing so will ultimately provide instrumental social value.

Procedural justice has also been associated with social identification. Because the police are highly visible representatives of the state and justice system, their treatment of citizens communicates inclusion or exclusion and can apprise people of their own position in society. Crucially, in Western countries, the police symbolise law-abiding citizenship and the country at large. Therefore, procedurally just policing evokes relational considerations and inspires shared group membership with these superordinate categories. Specifically, when people are treated with procedural fairness they are more likely to see themselves as being law-abiding citizens and belonging to the nation-state. This identification involves the feeling of having status within that particular group and that the existence of the group is worthwhile and something to cherish (Bradford 2014; Meares 2017; Radburn and Stott 2018).

Some scholars have argued that enhancing social identification can play a role in quelling uncertainty or anxiety especially for those whose ties to the superordinate groups are only tenuous, such as young people, immigrants, or ethnic minorities. Social identification, informed by procedural fairness (or lack thereof), helps individuals to monitor their place in the status hierarchy, thus providing them with feedback and reducing uncertainty (Bradford 2014; Colquitt et al. 2012; Murphy and Mazerolle 2018). It is plausible that both a sense of power and social identification contribute to reducing uncertainty.

Contrary to the broad agreement regarding the importance of social identification, there are two slightly different alternative theories on how procedural justice and social identification are related to each other. According to the group-value model (Lind and Tyler 1988; Tyler 1989), social identification precedes interactions with authorities. Encounters with the police are still important because just treatment

conveys identity-relevant information. However, this will be understood differently by people identifying with certain groups and thus the effects of procedural justice are moderated by one's particular identity. For instance, people who belong to an ethnic minority can find it more difficult to identify with a superordinate group compared to those who belong to the majority social group. Most studies have found mixed evidence regarding the group-value model, indicating that the varying effects of identity are idiosyncratic (i.e., highly dependent on a particular identity) (Bradford 2014; Murphy et al. 2017; Murphy, Sargeant, and Cherney 2015; Murphy and Mazerolle 2018).

By contrast, the group-engagement model (Blader and Tyler 2009; Tyler and Blader 2003) postulates that identity-relevant information passed on by procedurally just treatment shows individuals that they belong to the superordinate social group, which in turn encourages them to internalise the group's values. Hence, social identification mediates the impact of procedural justice on legitimacy and cooperation. In other words, procedurally fair treatment by the police reaffirms people's social position and identity, which subsequently makes them more likely to follow the rules and norms of the superordinate group, thus attributing a legitimate right-to-rule to the authorities. Again, the evidence base for this approach is mixed; some studies have found an indirect effect (Bradford, Hohl, et al. 2015; Bradford, Milani, et al. 2017; Bradford, Murphy, and Jackson 2014), whilst others have not (Bradford 2014).

In this thesis, Paper 2 is dedicated to identifying the causal pathways connecting procedural justice to legitimacy. As part of this enquiry, both personal sense of power and social identification are examined. In addition, Paper 1 includes social identity as an outcome variable and Paper 3 includes personal sense of power as one of the mediators of the impact of procedural justice on compliance and cooperation.

Causal evidence and procedural justice

Finally, I now briefly overview the accumulated causal evidence in the procedural justice policing literature. As mentioned earlier, a recent review and rejoinder have addressed this topic in great detail (Nagin and Telep 2017; Tyler 2017). I revisit the empirical evidence because the main aim of this thesis is to identify causal mechanisms for which this is especially pertinent. Also, since the two reviews, there have been several further contributions which deserve attention.

One of the main difficulties in surveying the causal evidence is that several criminal justice interventions have incorporated elements of procedural justice into them. For instance, a meta-analysis (Braga, Welsh, and Schnell 2015) of thirty experimental and quasi-experimental studies on the effectiveness of broken windows theory showed that the successful interventions were community-based, problem-solving oriented, and tended to apply– whether intentionally or unintentionally – procedurally just principles. By contrast, aggressive order maintenance tactics (e.g., misdemeanour arrests, ordinance violation summons) did not appear to have much of an effect. Procedurally just practices have been included among others in programmes such as violence-reduction strategy meetings (Papachristos and Kirk 2015), hotspot policing (Bennett, Newman, and Sydes 2017), offender notification forums (Wallace et al. 2016), and so on. Because in such instances it is impossible to disentangle the impact of procedural justice from other elements, hence these are not discussed here.

A further feature of this overview is that, unlike Nagin and Telep (2017), I only include studies where (1) procedural justice or police legitimacy was measured (instead of trust in the police), and (2) public perception was considered (instead of answers from trained coders or the police). Moreover, (3) I do not include longitudinal studies because none have used quasi-experimental methods (e.g., difference-in-differences) and the standard longitudinal methods require very strong identification assumptions for causal inference which – for most cases – are unlikely to hold true (Hamaker, Kuiper, and Grasman 2015; Imai and Kim 2016; Robins 1987). These criteria however limit the review to a handful of articles.

Nivette and Akoensi's (2017) vignette study of hypothetical citizen- and police-initiated encounters fielded in Accra, Ghana, found that when police effectiveness, lawfulness (corruption), and procedural justice were manipulated, a lack of effectiveness and lawfulness could undermine the impact of procedural justice on satisfaction with the police (legitimacy was not measured). A similar vignette study by Reisig, Mays, and Telep (2018) manipulated procedural justice and the outcome (i.e., receiving a citation) of the encounter. They found that people who were treated procedurally fairly were more likely to be satisfied with the outcome, comply with officers, and accept their decisions, regardless of the outcome.

There have been several studies where people were asked to watch procedurally just or unjust encounters and were subsequently asked for their opinions. Using a large and diverse sample from a panel, Braga et al. (2014) found that people's

perception of videotaped police encounters was influenced by what they were lead to believe about the social context (e.g., community relations) and also by the procedural justice of their most recent personal contact with the police. Lowrey et al.'s (2016) paper on university students showed that a videotaped procedurally just encounter had a positive effect on encounter-specific questions on duty to obey and willingness to cooperate but those effects were absent when general questions were asked about the police. Finally, Maguire et al.'s (2017) similar study on university students also found similar results, with a strong impact of procedural justice on the encounter-specific questions, but only a weak influence on the general police-related ones.

The treatment's distinct effects on encounter-specific opinions compared to general views is not exclusive to evaluating videotaped police stops. Sahin et al. (2017) conducted a field experiment in Adana, Turkey, where the police either incorporated procedurally just messages in their communications during traffic stops or carried on with their usual behaviour. People in the treatment group evaluated the specific encounter as more fair and legitimate, but those views did not carry over to the police in general. Sahin et al.'s (2017) study was inspired by the Queensland Community Engagement Trial (QCET), which found relatively strong evidence that procedurally just messaging increases both the perception of procedural justice and legitimacy, and has an impact on future cooperation as well (Mazerolle et al. 2013a). Another partial replication of QCET, the Scottish Community Engagement Trial produced the opposite results, with the supposedly procedurally just group reducing perceived procedural justice (MacQueen and Bradford 2015, 2017). As both Paper 1 and Paper 3 use the dataset of this study and because they include a major reassessment of the findings, I continue the discussion there.

This brief overview shows the limited causal evidence base of procedural justice policing. In the upcoming four papers, I estimate not only the average treatment effect – as all of the studies reviewed have – but also the mediated effects to tap into causal mechanisms.

Overview of the empirical component and research questions

Overview of the papers and the theoretical component

The substantive component of this thesis comprises three single-authored and one jointly-authored papers. This section briefly outlines these papers, chiefly with regards to the research questions and how they address the overall aims of the thesis. Each paper uses a different analytical technique from the causal mediation analysis family. This methodological component of the thesis is summarised separately at the end of this section.

Paper 1, *Prying Open the Black Box of Causality: A Causal Mediation Analysis Test of Procedural Justice Policing*, uses data from the Scottish Community Engagement Trial (ScotCET). This block-randomised controlled trial produced unexpected opposite effects where participants in the treatment group (i.e., those who met police officers asked to use procedurally just messages) reported significantly lower levels of perceived procedural justice (MacQueen and Bradford 2015). The paper starts by assessing this apparent failure of implementation (MacQueen and Bradford 2017), specifically by examining the selection bias, treatment consistency, and treatment effect and design heterogeneity. It finds that despite the surprising finding, the data can still be harnessed for certain goals, as the treatment effect was only attributable to the research design.

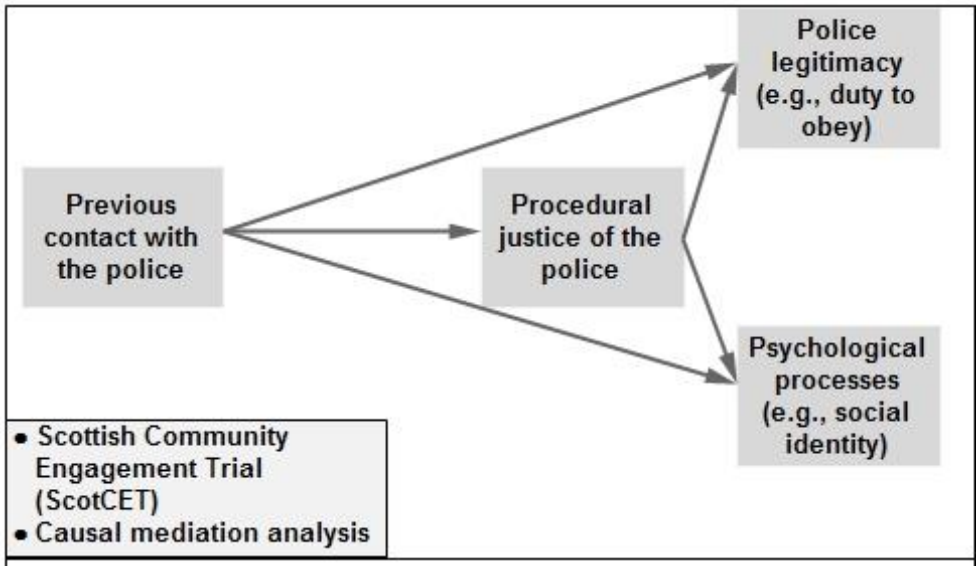


Figure 4 Outline of the theoretical model assessed in Paper 1

After these preliminary steps, the paper uses causal mediation analysis and corresponding sensitivity analysis techniques to assess this fundamental question of the procedural justice literature (Tyler 2006b):

> *Q1 Does procedural justice mediate the impact of previous contact on psychological processes and legitimacy?*

The model tested by Paper 1 is depicted in Figure 4. Q1 received a qualified yes. Procedural justice appears to fully mediate the impact of previous contact on normative alignment with the police and moderately strongly on duty to obey the police when average effects are considered. These pathways are supported by the results from the sensitivity analyses for unmeasured confounding. By contrast, procedural justice only channels a fraction of the causal effect towards social identification, which seems to be highly sensitive. This implies that while there is causal evidence that contact with the police influences police legitimacy through perceived procedural fairness, for social identification the causal mechanism is only tentative.

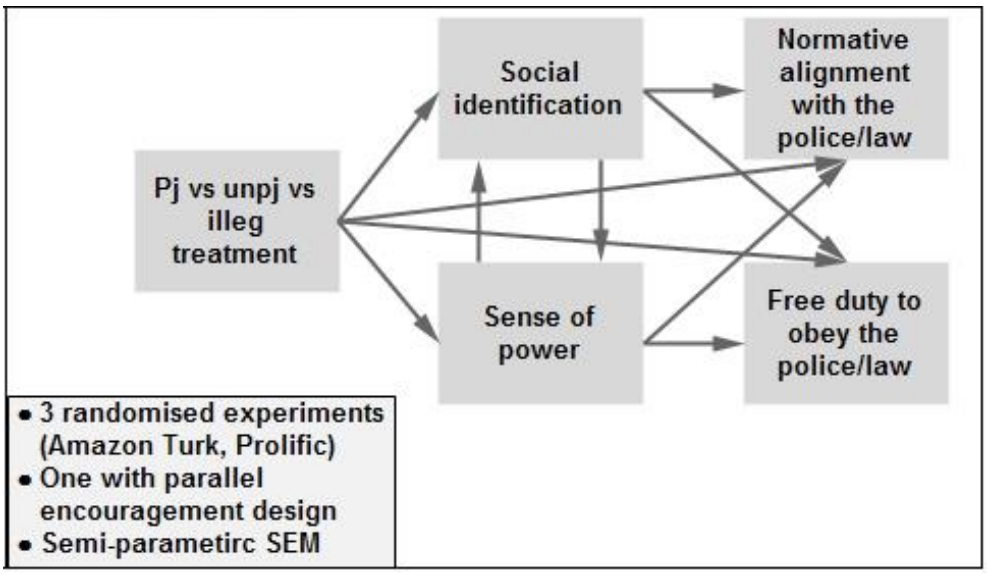


Figure 5 Outline of the theoretical model assessed in Paper 2

Paper 2, *"It's nice to be empowered" – An experimental assessment of psychological drivers of police legitimacy using statistical and design-based*

approaches estimating causally mediated effects, builds on the results of Paper 1. If procedural justice mediates the impact of contact on legitimacy then one should ask:

> *Q2 Which psychological mechanisms carry the impact of procedural justice on legitimacy?*

To answer Q2, I conducted three experiments (Figure 5). For Study 1 and Study 3, I crowdsourced participants from the United States (Amazon Turk), while for Study 2, I crowdsourced them from the UK (Prolific Academic). In each experiment three conditions were compared: a procedurally just, a procedurally unjust, and a breach of boundaries condition. The two prime candidates for mediating the impact of procedural justice were social identification (Bradford 2014; Bradford, Murphy, et al. 2014) and sense of power (Anderson, John, and Keltner 2012; Mentovich 2012), but in Study 1 police grip on power was also considered as an alternative mediator, and was complemented by self-control in Study 2. As indicated by the arrows in Figure 5, these mediators were all assumed to be affected by the treatment and to confound each other. The outcomes were measures of legitimacy of the police and the law (Huq et al. 2017).

Surprisingly, from all mediators in Study 1 and Study 2, only personal sense of power transmitted the effect of procedural justice and only on normative alignment with the police and the law. This prompted Study 3, which used a parallel (encouragement) design to estimate the mediating role of sense of power on the normative alignment aspects of legitimacy and arrived at largely similar effects. To answer Q2, Paper 2 indicates that procedural justice influences the perceived normative appropriateness of the police and the law by giving the individual an augmented sense of control, power, and autonomy in expected future encounters with the police.

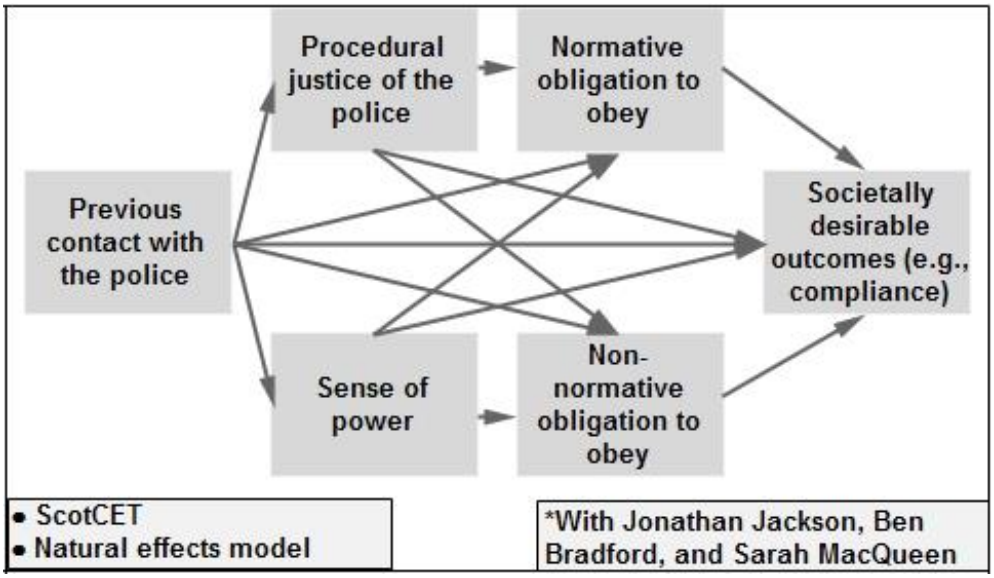


Figure 6 Outline of the theoretical model assessed in Paper 3

Paper 3, *"Truly Free Consent"? Clarifying the Nature of Legitimacy Using Causal Mediation Analysis*, was co-authored with Jonathan Jackson, Ben Bradford. and Sarah MacQueen. This paper revisits the ScotCET dataset with a model that tests all components of the comprehensive model outlined in the introduction. The goal of this paper was to differentiate between the normative and non-normative aspects of duty to obey (Bottoms and Tankebe 2012) and to explain how and why contact with the police affects cooperation with the police and compliance with the law. The research question asks:

> *Q3 Do procedural justice, sense of power, and police legitimacy mediate the impact of previous contact on compliance and cooperation?*

The tested model (Figure 6) found that the primary mediator of the impact of contact was normative duty to obey, but in the presence of non-normative obligation, both sense of power and procedural justice mediated the effect of contact on cooperation and compliance. Non-normative obligation did not transmit the effect of the contact on either of the outcomes and seems to exist outside of the model. Thus, to answer Q3, the findings imply that free duty to obey (i.e., police legitimacy) is the most important causal mechanism connecting contact with societally desirable outcomes.

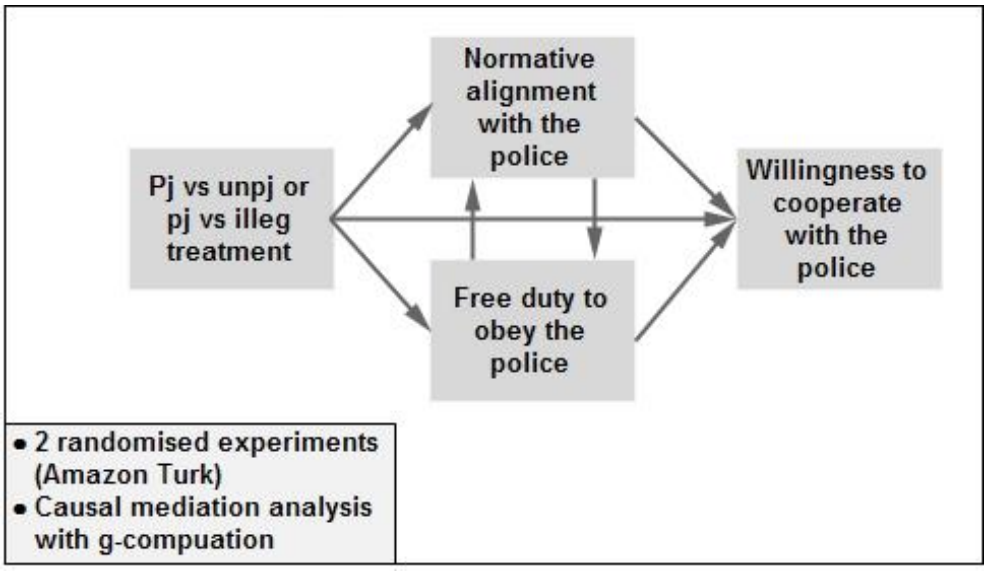


Figure 7 Outline of the theoretical model assessed in Paper 4

Paper 4, *Testing Complex Social Theories with Causal Mediation Analysis and G-Computation: Towards a Better Way to Do Causal Structural Equation Modelling*, scrutinises the discovery of Paper 3. If indeed legitimacy is the most important causal mechanism producing societally desirable outcomes then:

Q4 Does legitimacy mediate the impact of procedural justice on willingness to cooperate?

Compared to Paper 3, Paper 4 is more modest in its scope (Figure 7). In two randomised experiments with US participants (crowdsourced from Amazon Turk) procedural justice and respect for boundaries were manipulated, whilst police legitimacy and willingness to cooperate were measured. The two analytical approaches used in this paper arrived at very similar results, only finding evidence for the mediating role of normative alignment on cooperation. Thus, the answer to Q4 is that only the normative alignment aspect of legitimacy appears to mediate the effect of procedural justice on willingness to cooperate with the police.

Overview of the methodological component

Finally, I provide a brief overview from a methodological viewpoint; the details of each method is described in the paper where it is applied. In particular, Paper 1 and Paper 4 can be considered reviews of causal mediation analysis with single and

multiple mediators respectively. All methods used in this thesis are summarised by Table 1 with information regarding the way of estimation (statistical or design-based), the number of mediators handled, the relationship between the mediators, which paper they are used in, the method of estimation, the derived indirect effects, and the parametric assumptions required for the estimation. For all approaches, apart from the design-based estimation, a tailored version of the sequential ignorability assumption (Pearl 2001) must be satisfied for causal identification, which is discussed in each corresponding paper of each method.

Overall, causal mediation analysis is similar to other causal inference techniques in that using fewer assumptions is usually coupled with difficulties in estimation and/or decomposition and interpretability, whilst using more assumptions makes the effects easier to derive but less likely to be realistic. The methods can be categorised based on their number of mediators (single or multiple) and whether they relied on a statistical or design-based estimation. For statistical estimation, the causal identification assumptions need to be satisfied and an appropriate modelling strategy pursued, depending on the number of mediators and the causal structure. By contrast, design-based estimation is only possible if the data collection strategy is carried out according to specific requirements.

Starting with the statistical estimation, in case of a single mediator, the natural indirect effect (the causally mediated effect) is non-parametrically identifiable (Imai et al. 2011; Imai, Keele, and Tingley 2010). In case of multiple mediators that are not independent of each other, three approaches can be pursued. First, joint natural indirect effects can be estimated, which remain non-parametrically identifiable but do not allow for further decomposition (VanderWeele and Vansteelandt 2014). Alternatively, causal/sequential order can be assumed where causal mediators follow a pre-determined causal order. From the two approaches applied here, g-computation provides the finest decomposition, estimating the joint and separate effect of each mediator, but this requires adherence to strong parametric assumptions (Daniel et al. 2015). By contrast, natural effects models have fewer assumptions but they bundle up the jointly mediated effects and attribute them to the first mediator (natural indirect effect), only estimating partial indirect effects for every subsequent mediator (Steen et al. 2017). A third approach for multiple mediators is to assume post-treatment confounding (i.e., that the mediators are all affected by the treatment and have an impact on each other, but no causal order can be established). Here, g-computation has

more flexible parametric assumptions, but it becomes computationally demanding, the estimation taking days even with cluster-computing systems and relatively small datasets (De Stavola et al. 2015). By contrast, semi-parametric structural equation models allow easier estimation, but rely on more stringent assumptions (Imai and Yamamoto 2013).

Beyond these statistical approaches, design-based strategies are also viable to estimate natural indirect effects. This thesis used a parallel (encouragement) design, which is essentially a randomised experiment where half of the participants also randomly receive a second manipulation for the mediator (Imai, Tingley, and Yamamoto 2013). The main difference between the parallel and parallel encouragement design is the assumption regarding the second manipulation. The parallel design postulates that the second manipulation was perfect and thus effects are estimable for the whole population. By contrast, the parallel encouragement design assumes imperfection and only estimates the effects for the compliers. For the parallel design, parametric restrictions must be made to make the causal effects point-identified, whilst for the parallel encouragement design, only sharp bounds can be derived among the compliers separately for the treatment and control group.

		Paper	*Method of estimation*	*Decomposition*	*Parametric assumptions*
Single mediator		Paper 1	Semi-parametric causal mediation analysis	Natural Indirect Effect	Non-parametrically identifiable
Multiple mediators	*Joint effects*	Paper 3	Natural effects model	Joint Natural Indirect Effect	Non-parametrically identifiable
	Causal order/ sequential mediation			Natural Indirect Effect Partial Indirect Effect(s)	Identified through the imputation of weighted nested counterfactuals
	Post-treatment confounding	Paper 4	G-computation	(Summary) Natural Indirect Effects (Summary) Joint Effect	Linearity Influence of a sensitivity parameter
				Natural Indirect Effect	Linearity No-interaction (in expectation, two solutions)
		Paper 2	Semi-parametric structural equation modelling	Natural Indirect Effect	Linearity No-interaction (in expectation)
Design-based estimation			Parallel design	Natural Indirect Effect	No-interaction (for each unit).
				Natural Indirect Effect	Linearity No-interaction (in expectation)
				Natural Indirect Effect (control and treated)	Not point-identified
			Parallel encouragement design	Complier Natural Indirect Effect (control and treated)	Not point-identified

Table 1 A methodological overview of the empirical component of the thesis

References for the introduction and overview

Abulof, Uriel. 2016. "Public Political Thought: Bridging the Sociological: Philosophical Divide in the Study of Legitimacy." *British Journal of Sociology* 67(2):371–91.

Anderson, Cameron, John Angus D. Hildreth, and Laura Howland. 2015. "Is the Desire for Status a Fundamental Human Motive ? A Review of the Empirical Literature." *Psychological Bulletin* 141(3):574–601.

Anderson, Cameron, Oliver P. John, and Dacher Keltner. 2012. "The Personal Sense of Power." *Journal of Personality* 80(2):313–44.

Augustyn, Megan Bears. 2015. "The (Ir) Relevance of Procedural Justice in the Pathways to Crime." *Law and Human Behavior* 39(4):388.

Augustyn, Megan Bears. 2016. "Updating Perceptions of (In)Justice." *Journal of Research in Crime and Delinquency* 53(2):255–86.

Baker, T. et al. 2013. "Female Inmates' Procedural Justice Perceptions of the Police and Courts: Is There a Spill-Over of Police Effects?" *Criminal Justice and Behavior* 20(10):1–19.

Barclay, Laurie J., Michael R. Bashshur, and Marion Fortin. 2017. "Motivated Cognition and Fairness: Insights, Integration, and Creating a Path Forward." *Journal of Applied Psychology*. 102(6): 867-889.

Baron, Reuben M. and David A. Kenny. 1986. "Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6):173–82.

Becker, Gary. 1968. "Crime and Punishment - An Economic Approach." *Journal of Political Economy* 76(2):169–217.

Becker, Gary. 1974. "Crime and Punishment - A Critical Review (Revised Version)." Pp. 1–54 in *Essays in the Economics of Crime and Punishment*. Columbia University Press.

Beebee, Helen, Christopher Hitchcock, and Peter Menzies, eds. 2009. *The Oxford Handbook of Causation*. Oxford University Press.

Beetham, David. 1991. *The Legitimacy and Legitimation of Power*. London: ESRC.

Beetham, David. 2013. "Revisiting Legitimacy Twenty Years On." Pp. 19–36 in *Legitimacy and Criminal Justice - An International Exploration*, edited by J. Tankebe and A. Liebling. Oxford University Press.

Bell, Raoul and Axel Buchner. 2012. “How Adaptive Is Memory for Cheaters?” *Current Directions in Psychological Science* 21(6):403–8.

Bennett, Sarah, Mike Newman, and Michelle Sydes. 2017. “Mobile Police Community Office: A Vehicle for Reducing Crime, Crime Harm and Enhancing Police Legitimacy?” *Journal of Experimental Criminology* 13(3):1–12.

Bianchi, Emily et al. 2015. “Trust in Decision-Making Authorities Dictates the Form of the Interactive Relationship between Outcome Favorability and Procedural Fairness.” *Personality and Social Psychology Bulletin* 41(1):19–34.

Blader, Steven L. and Ya-Ru Chen. 2012. “Differentiating the Effects of Status and Power: A Justice Perspective.” *Journal of Personality and Social Psychology* 102(5):994–1014.

Blader, Steven L. and Tom R. Tyler. 2009. “Testing and Extending the Group Engagement Model: Linkages between Social Identity, Procedural Justice, Economic Outcomes, and Extrarole Behavior.” *Journal of Applied Psychology* 94(2):445–64.

Bonnefon, J., A. Hopfensitz, and W. De Neys. 2013. “The Modular Nature of Trustworthiness Detection.” *Journal of Experimental Psychology General* 142(1):143–50.

Van den Bos, K. 2001. “Uncertainty Management: The Influence of Uncertianty Salience on Reactions to Perceived Procedural Fairness.” *Journal of Personality and Social Psychology Social Psychology* 80(6):931–41.

Bottoms, Anthony and Justice Tankebe. 2012. “Beyond Procedural Justice: A Dialogic Approach To Legitimacy in Criminal Justice.” *Journal of Criminal Law & Criminology* 102(1):119–70.

Bradford, Ben. 2014. “Policing and Social Identity: Procedural Justice, Inclusion and Cooperation between Police and Public.” *Policing and Society* 24(1):22–43.

Bradford, Ben. 2017. *Stop and Search and Police Legitimacy*. Routledge.

Bradford, Ben, Katrin Hohl, Jonathan Jackson, and Sarah MacQueen. 2015. “Obeying the Rules of the Road.” *Journal of Contemporary Criminal Justice* 31(2):171–91.

Bradford, Ben, Aziz Huq, Jonathan Jackson, and Benjamin Roberts. 2014. “What Price Fairness When Security Is at Stake? Police Legitimacy in South Africa.” *Regulation and Governance* 8(2):246–68.

Bradford, Ben, Jonathan Jackson, and Mike Hough. 2017. “Ethnicity, Group Position and Police Legitimacy: Early Findings from the European Social Survey.” Pp.

73–95 in *Police-Citizen Relations - A Comparative Investigation of Sources and Impediments of Legitimacy around the World*, edited by S. Roche and D. Oberwitter. Routledge.

Bradford, Ben, Jonathan Jackson, and Elizabeth A. Stanko. 2009. "Contact and Confidence: Revisiting the Impact of Public Encounters with the Police." *Policing and Society* 19(1):20–46.

Bradford, Ben, Jenna Milani, and Jonathan Jackson. 2017. "Identity, Legitimacy and 'Making Sense' of Police Use of Force." *Policing: An International Journal of Police Strategies & Management* 40(3):614–27.

Bradford, Ben, Kristina Murphy, and Jonathan Jackson. 2014. "Officers as Mirrors." *British Journal of Criminology* 54(4):527–50.

Bradford, Ben, Elise Sargeant, Kristina Murphy, and Jonathan Jackson. 2015. "A Leap of Faith? Trust in the Police Among Immigrants in England and Wales." *British Journal of Criminology* (December 2015):1–21. Retrieved (http://bjc.oxfordjournals.org/content/early/2015/12/30/bjc.azv126.abstract).

Bradford, Ben, Elizabeth Stanko, and Jonathan Jackson. 2012. *Just Authority? - Trust in the Police in England and Wales*. Routledge.

Braga, Anthony A., Brandon C. Welsh, and Cory Schnell. 2015. "Can Policing Disorder Reduce Crime? A Systematic Review and Meta-Analysis." *Journal of Research in Crime and Delinquency* 52(4):567–88.

Braga, Anthony a., Christopher Winship, Tom R. Tyler, Jeffrey Fagan, and Tracey L. Meares. 2014. "The Salience of Social Contextual Factors in Appraisals of Police Interactions with Citizens: A Randomized Factorial Experiment." *Journal of Quantitative Criminology* 30(4):599–627.

Brayne, S. 2014. "Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment." *American Sociological Review* 79(3):367–91.

Brooks Holliday, S., C. King, and K. Heilbrun. 2013. "Offenders' Perceptions of Risk Factors for Self and Others: Theoretical Importance and Some Empirical Data." *Criminal Justice and Behavior* 40(9):1044–61.

Cavanagh, Caitlin and Elizabeth Cauffman. 2017. "What They Don't Know Can Hurt Them: Mothers' Legal Knowledge and Youth Re-Offending." *Psychology, Public Policy, and Law* 23(2):141–53.

Charles, Kerwin Kofi and Steven N. Durlauf. 2013. "Pitfalls in the Use of Time Series Methods to Study Deterrence and Capital Punishment." *Journal of Quantitative*

Criminology 29(1):45–66.

Cherney, a. and K. Murphy. 2013. "Policing Terrorism with Procedural Justice: The Role of Police Legitimacy and Law Legitimacy." *Australian & New Zealand Journal of Criminology* 46(3):403–21.

Colquitt, Jason A., Jeffrey A. LePine, Ronald F. Piccolo, Cindy P. Zapata, and Bruce L. Rich. 2012. "Explaining the Justice-Performance Relationship: Trust as Exchange Deepener or Trust as Uncertainty Reducer?" *Journal of Applied Psychology* 97(1):1–15.

Daniel, R. M., B. L. De Stavola, S. N. Cousens, and S. Vansteelandt. 2015. "Causal Mediation Analysis with Multiple Mediators." *Biometrics* 71(1):1–14.

Desmond, M., A. V. Papachristos, and D. S. Kirk. 2016. "Police Violence and Citizen Crime Reporting in the Black Community." *American Sociological Review* 81(5):857–76.

Durlauf, Steven N. and Daniel S. Nagin. 2011. "Imprisonment and Crime: Can Both Be Reduced?" *Criminology & Public Policy* 10(1):13–54.

Epp, Charles R., Steven Maynard-Moody, and Donald P. Haider-Markel. 2014. *Pulled Over: How Police Stops Define Race and Citizenship*. University of Chicago Press.

Fairfield, Tasha and Andrew Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." *Political Analysis* 25(3):363–80.

Foucault, Michel. 1977. *Discipline and Punish: The Birth of the Prison*. Random House.

Gau, Jacinta M. 2013. "Consent Searches as a Threat to Procedural Justice and Police Legitimacy: An Analysis of Consent Requests During Traffic Stops." *Criminal Justice Policy Review* 24(6):759–77.

Gau, Jacinta M. and Rod K. Brunson. 2010. "Procedural Justice and Order Maintenance Policing: A Study of Inner-City Young Men's Perceptions of Police Legitimacy." *Justice Quarterly* 27(2):255–79.

Gau, Jacinta M. and Rod K. Brunson. 2012. "'One Question Before You Get Gone...' Consent Search Requests as a Threat to Perceived Stop Legitimacy." *Race and Justice* 2(4):250–73.

Gauthier, Jane Florence, Lisa M. Graziano, Jane Florence Gauthier, and Lisa M. Graziano. 2018. "News Media Consumption and Attitudes about Police : In

Search of Theoretical Orientation and Advancement of Theoretical Orientation and Advancement." *Journal of Crime and Justice* In press:1–17. Retrieved (https://doi.org/10.1080/0735648X.2018.1472625).

Goffman, a. 2009. "On the Run: Wanted Men in a Philadelphia Ghetto." *American Sociological Review* 74(3):339–57.

Goldstein, Daniel M. 2003. "'In Our Own Hands': Lynching, Justice, and the Law in Bolivia." *American Ethnologist* 30(1):22–43.

Haberman, Cory P. 2016. "A View inside the 'Black Box' of Hot Spots Policing from a Sample of Police Commanders." *Police Quarterly* 19(4):488–517.

Hamaker, Ellen L., Rebecca M. Kuiper, and Raoul P. P. P. Grasman. 2015. "A Critique of the Cross-Lagged Panel Model." *Psychological Methods* 20(1):102–16.

Hamm, J. A., R. Trinkner, and J. D. Carr. 2017. "Fair Process, Trust, and Cooperation: Moving Toward an Integrated Framework of Police Legitimacy." *Criminal Justice and Behavior* 44(9):1183–1212.

Hedström, Peter and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36(1):49–67.

Heider, Fritz. 1958. *The Psychology of Interpersonal Relations*. Wiley.

Higginson, Angela and Lorraine Mazerolle. 2014. "Legitimacy Policing of Places: The Impact of Crime and Disorder." *Journal of Experimental Criminology* 10(4):429–57.

Hinsch, Wilfried. 2010. "Justice, Legitimacy, and Constitutional Rights." *Critical Review of International Social and Political Philosophy* 13(1):39–54.

Von Hippel, William, Jessica L. Lakin, and Richard J. Shakarchi. 2005. "Individual Differences in Motivated Social Cognition - The Case of Self-Serving Information Processing." *Personality and Social Psychology Bulletin* 31(10):1347–57.

Hohl, Katrin, Ben Bradford, and Elizabeth a. Stanko. 2010. "Influencing Trust and Confidence in the London Metropolitan Police: Results from an Experiment Testing the Effect of Leaflet Drops on Public Opinion." *British Journal of Criminology* 50(3):491–513.

Hough, Mike. 2012. "Researching Trust in the Police and Trust in Justice: A UK Perspective." *Policing and Society* 22(3):332–45.

Hough, Mike, Jonathan Jackson, and Ben Bradford. 2013. "Legitimacy, Trust and Compliance: An Empirical Test of Procedural Justice Theory Using the European

Social Survey." Pp. 326–53 in *Legitimacy and Criminal Justice - An International Exploration*, edited by J. Tankebe and A. Liebling. Oxford University Press.

Huq, A., T. Tyler, and S. Schulhofer. 2011. "Mechanisms for Eliciting Cooperation in Counter Terrorism Policing: Evidence from the United Kingdom." *Journal of Empirical Legal Studies* 8(4):728–61.

Huq, A. Z. Aziz H., J. Jackson, and R. J. Trinker. 2017. "Legitimating Practices: Revisiting the Predicates of Police Legitimacy." *British Journal of Criminology* (57):1101–22.

Imai, K., D. Tingley, and T. Yamamoto. 2013. "Experimental Designs for Identifying Causal Mechanisms (with {D}iscussion)." *J. Roy. Stat. Soc., A* 176:5–51.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological Methods* 15(4):309–34. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/20954780).

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105(4):765–89.

Imai, Kosuke and In Song Kim. 2016. "When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *Unpublished*. Retrieved (http://imai.princeton.edu/research/files/FEmatch.pdf).

Imai, Kosuke and Teppei Yamamoto. 2013. "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments." *Political Analysis* 21(2):141–71.

Intravia, Jonathan, Eric A. Stewart, Patricia Y. Warren, and Kevin T. Wolff. 2016. "Neighborhood Disorder and Generalized Trust: A Multilevel Mediation Examination of Social Mechanisms." *Journal of Criminal Justice* 46:148–58.

Jackson, Jonathan et al. 2012. "Why Do People Comply with the Law?" *British Journal of Criminology* 52(6):1051–71.

Jackson, Jonathan. 2018. "Norms, Normativity, and the Legitimacy of Justice Institutions: International Perspectives." *Annual Reveiw of Law and Social Sciences* 14 In pres.

Jackson, Jonathan, Muhammad Asif, Ben Bradford, and Muhammad Zakria Zakar. 2014. "Corruption and Police Legitimacy in Lahore, Pakistan." *British Journal of Criminology* 54(6):1067–88.

Jackson, Jonathan and Ben Bradford. 2009. “Crime, Policing and Social Order: On the Expressive Nature of Public Confidence in Policing.” *British Journal of Sociology* 60(3):493–521.

Jackson, Jonathan, Ben Bradford, Ian Brunton-smith, and Emily Gray. 2018. “In the Eye of the (Motivated) Beholder: Towards a Motivated Cognition Perspective on Disorder Perceptions.” Pp. 253–71 in *The Routledge International Handbook on Fear of Crime*, edited by M. Lee and G. Mythen. Routledge.

Jackson, Jonathan and Jacinta M. Gau. 2015. “Carving up Concepts? - Differentiating between Trust and Legitimacy in Public Attitudes towards Legal Authority.” Pp. 49–69 in *Interdiisciplinary Perspectives on Trust - Towards Theoretical and Methodological Integration*, edited by E. Shockley, T. M. S. Neal, L. PytlikZillig, and B. Bornstein. Springer.

Jackson, Jonathan, Aziz Z. Huq, Ben Bradford, and Tom R. Tyler. 2013. “Monopolizing Force? Police Legitimacy and Public Attitudes toward the Acceptability of Violence.” *Psychology, Public Policy, and Law* 19(4):479–97.

Johnson, Devon, Edward R. Maguire, and Joseph B. Kuhns. 2014. “Public Perceptions of the Legitimacy of the Law and Legal Authorities: Evidence from the Caribbean.” 48(4):947–78.

Johnson, R. Burke, Federica Russo, and Judith Schoonenboom. 2017. “Causation in Mixed Methods Research: The Meeting of Philosophy, Science, and Practice.” *Journal of Mixed Methods Research* In press:1–20.

Jost, John T., Jack Glaser, Arie W. Kruglanski, and Frank J. Sulloway. 2003. “Political Conservatism as Motivated Social Cognition.” *Psychological Bulletin* 129(3):339–75.

Kahneman, Daniel. 2012. *Thinking Fast and Slow*. Penguin.

Keele, Luke. 2015. “Causal Mediation Analysis Warning! Assumptions Ahead.” *American Journal of Evaluation* 46(4):500–513.

Kirk, David S. and Andrew V. Papachristos. 2011. “Cultural Mechanisms and the Persistence of Neighborhood Violence.” *American Journal of Sociology* 116(4):1190–1233.

Kirk, David S. and Sara Wakefield. 2018. “Collateral Consequences of Punishment: A Critical Review and Path Forward.” *Annual Review of Criminology* 1(9):1–24.

Kleck, G. and J. C. Barnes. 2008. “Deterrence and Macro-Level Perceptions of Punishment Risks: Is There a ‘Collective Wisdom’?” *Crime & Delinquency*

59(7):1006–35.

Kochel, Tammy Rinehart. 2018. “Police Legitimacy and Resident Cooperation in Crime Hotspots: Effects of Victimisation Risk and Collective Efficacy.” *Policing and Society* 28(3):251–70.

Kruglanski, Arie W. 1996. “Motivated Social Cognition - Principles of the Interface.” Pp. 493–520 in *Social Psychology: Handbook of Basic Principles,* edited by E. Higgins and A. W. Kruglanski. Guilford Press.

van Lier, Jens, Russell Revlin, and Wim de Neys. 2013. “Detecting Cheaters without Thinking: Testing the Automaticity of the Cheater Detection Module.” *PLoS ONE* 8(1).

Lind, Allan E. 2001. “Fairness Heuristic Theory - Justice Judgments as Pivotal Cognitions in Organizational Relations.” Pp. 56–88 in *Advances in Organizational Justice,* edited by J. Greenberg and R. Cropanzano. New Lexington Press.

Lind, Allan and Tom Tyler. 1988. *The Social Psychology of Procedural Justice*. Springer.

Loader, Ian. 2006. “Policing, Recognition, and Belonging.” *Annals of the American Academy of Political and Social Science* 605(1):201–21.

Loader, Ian. 2014. “Why Do the Police Matter? Beyond the Myth of Crime-Fighting.” Pp. 52–63 in *The Future of Policing,* edited by Jennifer M. Brown.New York: Routledge.

Lowrey, Belén V., Edward R. Maguire, and Richard R. Bennett. 2016. “Testing the Effects of Procedural Justice and Overaccommodation in Traffic Stops: A Randomized Experiment.” *Criminal Justice and Behavior* 43(10):1430–49.

MacQueen, Sarah and Ben Bradford. 2015. “Enhancing Public Trust and Police Legitimacy during Road Traffic Encounters: Results from a Randomised Controlled Trial in Scotland.” *Journal of Experimental Criminology* 11(3):419–43.

MacQueen, Sarah and Ben Bradford. 2017. “Where Did It All Go Wrong? Implementation Failure—and More—in a Field Experiment of Procedural Justice Policing.” *Journal of Experimental Criminology* 13(3):321–45.

Madon, Natasha S., Kristina Murphy, and Adrian Cherney. 2017. “Promoting Community Collaboration in Counterterrorism: Do Social Identities and Perceptions of Legitimacy Mediate Reactions to Procedural Justice Policing?”

British Journal of Criminology 57(5):1114–64.

Maguire, Edward R., Belén V. Lowrey, and Devon Johnson. 2017. "Evaluating the Relative Impact of Positive and Negative Encounters with Police: A Randomized Experiment." *Journal of Experimental Criminology* 13(3):367–91.

Matsueda, Ross L. 2017. "Toward an Analytical Criminology: The Micro–Macro Problem, Causal Mechanisms, and Public Policy." *Criminology* 55(3):493–519.

Mazerolle, Lorraine, Emma Antrobus, Sarah Bennett, and Tom R. Tyler. 2013a. "Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice." *Criminology* 51(1):33–63.

Mazerolle, Lorraine, Sarah Bennett, Jacqueline Davis, Elise Sargeant, and Matthew Manning. 2013b. "Procedural Justice and Police Legitimacy: A Systematic Review of the Research Evidence." *Journal of Experimental Criminology* 9(3):245–74.

McCluskey, John. D. 2003. *Police Requests for Compliance - Coercive and Procedurally Just Tactics*. New York: LFB Scholarly Publishing.

Meares, Tracey. 2017. "Policing and Procedural Justice: Shaping Citizens' Identities to Increase Democratic Participation." *Northwestern University Law Review* 111(6):1525–35.

Mentovich, Avital. 2012. *The Power of Fair Procedures - The Effect of Procedural Justice on Perceptions of Power and Hierarchy*. Doctoral Thesis, New York University.

Moravcová, Eva. 2016. "Willingness to Cooperate with the Police in Four Central European Countries." *European Journal on Criminal Policy and Research* 22(1):171–87.

Morgan, Stephen L. and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. Cambridge University Press.

Muir, William K. 1977. *Streetcorner Politicians*. University of Chicago Press.

Murphy, Kristina and Adrian Cherney. 2012. "Understanding Cooperation with Police in a Diverse Society." *British Journal of Criminology* 52(1):181–201.

Murphy, Kristina, Adrian Cherney, and Marcus Teston. 2018. "Promoting Muslims' Willingness to Report Terror Threats to Police: Testing Competing Theories of Procedural Justice." *Justice Quarterly* In press:1–26.

Murphy, Kristina, Robert J. Cramer, Kevin A. Waymire, and Julie Barkworth. 2017.

"Police Bias, Social Identity, and Minority Groups: A Social Psychological Understanding of Cooperation with Police." *Justice Quarterly* In press:1–26. Retrieved (http://doi.org/10.1080/07418825.2017.1357742).

Murphy, Kristina and Lorraine Mazerolle. 2018. "Policing Immigrants : Using a Randomized Control Trial of Procedural Justice Policing to Promote Trust and Cooperation." *Australian & New Zealand Journal of Criminology* 51(1): 3–22.

Murphy, Kristina, Elise Sargeant, and Adrian Cherney. 2015. "The Importance of Procedural Justice and Police Performance in Shaping Intentions to Cooperate with the Police: Does Social Identity Matter?" *European Journal of Criminology* 12(6):719–38.

Murphy, Kristina, Tom R. Tyler, and Amy Curtis. 2009. "Nurturing Regulatory Compliance: Is Procedural Justice Effective When People Question the Legitimacy of the Law?" *Regulation and Governance* 3(1):1–26.

Myhill, Andy and Ben Bradford. 2012. "Can Police Enhance Public Confidence by Improving Quality of Service? Results from Two Surveys in England and Wales." *Policing and Society* 22(4):397–425.

Nagin, Daniel S. 2013. "Deterrence: A Review of the Evidence by a Criminologist for Economists." *Annual Review of Economics* 5(1):83–105. Retrieved (http://www.annualreviews.org/doi/abs/10.1146/annurev-economics-072412-131310).

Nagin, Daniel S. and Cody W. Telep. 2017. "Procedural Justice and Legal Compliance." *Annual Review of Law and Social Science* 13(1):5–28.

Nivette, Amy E. 2013. "Legitimacy and Crime: Theorizing the Role of the State in Cross-National Criminological Theory." *Theoretical Criminology* 18(1):93–111.

Nivette, Amy E. 2016. "Institutional Ineffectiveness, Illegitimacy, and Public Support for Vigilantism in Latin America." *Criminology* 54(1):142–75.

Nivette, Amy E. and Thomas D. Akoensi. 2017. "Determinants of Satisfaction with Police in a Developing Country: A Randomised Vignette Study." *Policing and Society* In press: 1–17. Retrieved (https://www.tandfonline.com/doi/full/10.1080/10439463.2017.1380643).

Ogburn, Elizabeth L., Oleg Sofrygin, Ivan Diaz, and Mark J. van der Laan. 2017. "Causal Inference for Social Network Data." *Unpublished Manuscript* 1–43. Retrieved (http://arxiv.org/abs/1705.08527).

Papachristos et al. 2012. "Why Do Criminals Obey the Law? The Influence of

Legitimacy and Social Networks on Active Gun Offenders." *The Journal of Crimibal Law & Criminologogy* 102(2):397–440.

Papachristos, Andrew V. and David S. Kirk. 2015. "Changing the Street Dynamic: Evaluating Chicago's Group Violence Reduction Strategy Papachristos and Kirk Evaluating Chicago's Group Violence Reduction Strategy." *Criminology and Public Policy* 14(3):525–58.

Papachristos, Andrew V., Christopher Wildeman, and Elizabeth Roberto. 2015. "Tragic, but Not Random: The Social Contagion of Nonfatal Gunshot Injuries." *Social Science and Medicine* 125:139–50.

Paternoster, Raymond. 2010. "How Much Do We Really Know About Criminal Deterrence?" *The Journal of Criminal Law and Criminology* 100(3):765–824.

Pearl, Judea. 2001. "Direct and Indirect Effects." *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* 411–20.

Penner, Erika K., Jodi L. Viljoen, Kevin S. Douglas, and Ronald Roesch. 2014. "Procedural Justice versus Risk Factors for Offending: Predicting Recidivism in Youth." *Law and Human Behavior* 38(3):225–37.

Pratt, Travis C., Francis T. Cullen, John Paul Wright, and K. R. Blevins. 2008. "The Empirical Status of Deterrence Theory: A Meta-Analysis." Pp. 367–96 in *Taking Stock: The Status of Criminological Theory*, edited by F. T. Cullen, J. P. Wright, and K. R. Blevins. Transaction Publishers.

van Prooijen, Jan-Willem, Kees van den Bos, and Henk a M. Wilke. 2002. "Procedural Justice and Status: Status Salience as Antecedent of Procedural Fairness Effects." *Journal of Personality and Social Psychology* 83(6):1353–61.

van Prooijen, Jan Willem. 2009. "Procedural Justice as Autonomy Regulation." *Journal of Personality and Social Psychology* 96(6):1166–80.

Proudfoot, Devon and Allan E. Lind. 2015. "Fairness Heuristic Theory, the Uncertainty Management Model, and Fairness at Work." Pp. 371–85 in *Oxford Handbook of Organizational Justice*, edited by R. Cropanzano and M. Ambrose. Oxford University Press.

Punch, Maurice. 2007. *Zero Tolerance Policing*. Policy Press.

Radburn, Matthew and Clifford Stott. 2018. "The Social Psychological Processes of ' Procedural Justice ': Concepts , Critiques and Opportunities." *Criminology and Criminal Justice* In Press. Retrieved (http://journals.sagepub.com/doi/abs/10.1177/1748895818780200).

Rawls, John. 1999. *A Theory of Justice: Revised Edition*. Harvard University Press.

Raz, Joseph. 2009. *Between Authority and Interpretation - On the Theory of Law and Practical Reason*. Oxford University Press.

Reiner, Robert. 2010. *The Politics of the Police - Third edition*. Oxford University Press.

Reiner, Robert. 2012. *In Praise of Fire Brigade Policing: Challenging the Police Role*.

Reisig, M. D. and C. Lloyd. 2008. "Procedural Justice, Police Legitimacy, and Helping the Police Fight Crime: Results From a Survey of Jamaican Adolescents." *Police Quarterly* 12(1):42–62.

Reisig, Michael D., Ryan D. Mays, and Cody W. Telep. 2018. "The Effects of Procedural Injustice during Police–citizen Encounters: A Factorial Vignette Study." *Journal of Experimental Criminology* 14(1):49–58.

Robins, James M. 1987. "A Graphical Approach to the Identification and Estimation of Causal Parameters in Mortality Studies with Sustained Exposure Periods." *Journal of Chronic Diseases* 40(2):139–61.

Rosenbaum, D. P. 2005. "Attitudes Toward the Police: The Effects of Direct and Vicarious Experience." *Police Quarterly* 8(3):343–65.

Sahin, Nusret, Anthony A. Braga, Robert Apel, and Rod K. Brunson. 2017. "The Impact of Procedurally-Just Policing on Citizen Perceptions of Police During Traffic Stops: The Adana Randomized Controlled Trial." *Journal of Quantitative Criminology* 33(4):701–26.

Sampson, Robert J., Christopher Winship, and Carly Knight. 2013. "Translating Causal Claims: Principles and Strategies for Policy-Relevant Criminology." *Criminology & Public Policy* 12(4):587–616.

Saylor, Ryan. 2018. "Why Causal Mechanisms and Process Tracing Should Alter Case Selection Guidance." *Sociological Methods & Research* In Press. Retrieved (http://journals.sagepub.com/doi/10.1177/0049124118769109).

Sindall, K., D. J. McCarthy, and I. Brunton-Smith. 2016. "Young People and the Formation of Attitudes towards the Police." *European Journal of Criminology* 14(3):344–64.

Skogan, Wesley G. 2006. *Police and Community in Chicago: A Tale of Three Cities*. Oxford University Press.

Slocum, Lee Ann, Stephanie Ann Wiley, and Finn-Aage Esbensen. 2016. "The Importance of Being Satisfied." *Criminal Justice and Behavior* 43(1):7–26.

De Stavola, Bianca L., Rhian M. Daniel, George B. Ploubidis, and Nadia Micali. 2015. "Mediation Analysis with Intermediate Confounding: Structural Equation Modeling Viewed through the Causal Inference Lens." *American Journal of Epidemiology* 181(1):64–80.

Steen, Johan, Tom Loeys, Beatrijs Moerkerke, and Johan Steen. 2017. "Flexible Mediation Analysis with Multiple Mediators." *American Journal of Epidemiology* 186(2):184–93.

Sunshine, Jason and Tom R. Tyler. 2003. "The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing." *Law and Society Review* 37(3):513–48.

Tabibnia, Golnaz, Ajay B. Satpute, and Matthew D. Lieberman. 2008. "The Sunny Side of Fairness." *Psychological Science* 19(4):339–47.

Tankebe, Justice. 2009. "Public Cooperation with the Police in Ghana: Does Procedural Fairness Matter?" *Criminology* 47(4):1265–93.

Tankebe, Justice. 2010. "Public Confidence in the Police: Testing the Effects of Public Experiences of Police Corruption in Ghana." *British Journal of Criminology* 50(2):296–319.

Tankebe, Justice. 2013. "Viewing Things Differently: The Dimensions of Public Perceptions of Police Legitimacy." *Criminology* 51(1):103–35.

Tankebe, Justice, Michael D. Reisig, Xia Wang, Michael D. Reisig, and Xia Wang. 2016. "Law and Human Behavior A Multidimensional Model of Police Legitimacy : A Cross-Cultural Assessment A Multidimensional Model of Police Legitimacy :" *Law and Human Behavior* 40(1):11–20.

Thibaut, John and Laurens Walker. 1975. *Procedural Justice: A Psychological Analysis*. Lawrence Erlbaum Associates.

Trinkner, Rick, Jonathan Jackson, and Tom R. Tyler. 2017. "Bounded Authority: Expanding 'Appropriate' Police Behavior Beyond Procedural Justice." *Law and Human Beh* 42(3):280–93.

Trinkner, Rick and Tom R. Tyler. 2016. "Legal Socialization : Coercion versus Consent in an Era of Mistrust." *Annual Review of Law and Social Science* 12:417–39.

Tyler, Phillip Atiba Goff, and Robert J. MacCoun. 2015. "The Impact of Psychological Science on Policing in the United States: Procedural Justice, Legitimacy, and Effective Law Enforcement." *Psychological Science in the Public Interest*

16(3):75–109.

Tyler, T., J. Fagan, and A. Geller. 2014. “Street Stops Police Legitimacy: Teachable Moments in Young Urban Men’s Legal Socialization.” *Journal of Empirical Legal Studies* 11(14):751–85.

Tyler, T. R. and Y. J. Huo. 2002. *Trust in the Law - Encouraging Public Cooperation With the Police and the Courts*. New York: Russell Sage Foundation.

Tyler, Tom. 1989. “The Psychology of Procedural Justice: A Test of the Group-Value Model.” *Journal of Personality and Social Psychology* 57(5):830–38.

Tyler, Tom and Jeffrey Fagan. 2008. “Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities?” *Ohio State Journal of Criminal Law* 6:231–75.

Tyler, Tom R. 2006a. “Psychological Perspectives on Legitimacy and Legitimation.” *Annual Review of Psychology* 57:375–400.

Tyler, Tom R. 2006b. *Why People Obey the Law*. Princeton: Princeton University Press.

Tyler, Tom R. 2011. *Why People Cooperate*. Princeton University Press.

Tyler, Tom R. 2017. “Procedural Justice and Policing: A Rush to Judgment?” *Annual Review of Law and Social Science* 13:29–53.

Tyler, Tom R. and Steven L. Blader. 2003. “The Group Engagement Model: Procedural Justice, Social Identity, and Cooperative Behavior.” *Personality and Social Psychology Review* 7(4):349–61.

Tyler, Tom R. and Jonathan Jackson. 2013. “Future Challenges in the Study of Legitimacy and Criminal Justice.” Pp. 83–104 in *Legitimacy and Criminal Justice - An International Exploration*, edited by J. Tankebe and A. Liebling. Wiley.

Tyler, Tom R. and Jonathan Jackson. 2014. “Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation, and Engagement.” *Psychology, Public Policy, and Law* 20(1):78–95.

Tyler, Tom R., Jonathan Jackson, and Avital Mentovich. 2015. “The Consequences of Being an Object of Suspicion - Potential Pitfalls of Proactive Police Contact.” *Journal of Empirical Legal Studies* 12(4):602–36.

Tyler, Tom R. and Allan E. Lind. 1992. “A Relational Model of Authority in Groups.” *Advances in Experimental Social Psychology* 25:115–91.

Tyler, Tom R. and Cheryl J. Wakslak. 2004. “Profiling and Police Legitimacy:

Procedural Justice Attributions of Motive." *Criminology* 42(2):253–81.

VanderWeele, Tyler J. 2015. *Explanation in Causal Inference - Methods for Mediation and Interaction*. Oxford University Press.

VanderWeele, Tyler J., Elizabeth L. Ogburn, and Eric J. Tchetgen Tchetgen. 2012. "Why and When 'Flawed' Social Network Analyses Still Yield Valid Tests of No Contagion." *Statistics, Politics, and Policy* 3(1):2151–63. Retrieved (http://www.degruyter.com/view/j/spp.2012.3.issue-1/2151-7509.1050/2151-7509.1050.xml).

VanderWeele, Tyler J. and Stijn Vansteelandt. 2014. "Mediation Analysis with Multiple Mediators." *Epidemiologic Methods* 2(1):95–115.

Wacquant, Loic. 2009. *Punishing the Poor: The Neoliberal Government of Social Insecurity*. Duke University Press.

Waddington, P. A. J., Kate Williams, Martin Wright, and Tim Newburn. 2015. "Dissension in Public Evaluations of the Police." *Policing and Society* 25(2):212–35.

Wallace, Danielle, Andrew V. Papachristos, Tracey Meares, and Jeffrey Fagan. 2016. "Desistance and Legitimacy: The Impact of Offender Notification Meetings on Recidivism among High Risk Offenders." *Justice Quarterly* 33(7):1237–64.

Van De Walle, Steven. 2009. "Confidence in the Criminal Justice System: Does Experience Count?" *British Journal of Criminology* 49(3):384–98.

Weber, Max. 1998. *Tanulmányok (Studies)*. Osiris.

Weller, Nicholas and Jeb Barnes. 2016. "Pathway Analysis and the Search for Causal Mechanisms." *Sociological Methods & Research* 45(3):424–57.

Wiley, Stephanie Ann, Lee Ann Slocum, and Finn Aage Esbensen. 2013. "The Unintended Consequences of Being Stopped or Arrested: An Exploration of the Labeling Mechanisms through Which Police Contact Leads to Subsequent Delinquency." *Criminology* 51(4):927–66.

Wilkinson, D. L., C. C. Beaty, and R. M. Lurry. 2009. "Youth Violence-- Crime or Self-Help? Marginalized Urban Males' Perspectives on the Limited Efficacy of the Criminal Justice System to Stop Youth Violence." *The ANNALS of the American Academy of Political and Social Science* 623(1):25–38.

Zimbardo, Philip. 2007. *The Lucifer Effect: Understanding How Good People Turn Evil*. Random House.

Zizumbo-Colunga, Daniel. 2017. "Community, Authorities, and Support for

Vigilantism: Experimental Evidence.” *Political Behavior* 39(4):1–27.

Paper 1: Prying Open the Black Box of Causality: A Causal Mediation Analysis Test of Procedural Justice Policing

Krisztián Pósch

Abstract

Objectives: Review causal mediation analysis as a method for estimating and assessing direct and indirect effects. Re-examine a field experiment with an apparent implementation failure. Test procedural justice theory by examining to which extent procedural justice mediates the impact of contact with the police on police legitimacy and social identity.

Methods: Data from a block-randomised controlled trial of procedural justice policing (the Scottish Community Engagement Trial) were analysed. All constructs were measured using surveys distributed during roadside police checks. Treatment implementation was assessed by analysing the treatment effect's consistency and heterogeneity. Causal mediation and sensitivity analysis were used to assess the mediating role of procedural justice.

Results: First, the treatment effect was fairly consistent and homogeneous, indicating that the treatment's effect is attributable to the design. Second, there is evidence that procedural justice channels the treatment's effect towards normative alignment (NIE=-0.207), duty to obey (NIE=-0.153), and social identity (NIE=-0.052), all of which are moderately robust to unmeasured confounding (ρ=0.3-0.6, LOVE=0.5-0.7).

Conclusions: The effect's consistency and homogeneity should be examined in future block-randomised designs. Causal mediation analysis is a versatile tool that can salvage experiments with systematic yet ambiguous treatment effects by allowing researchers to "pry open" the black box of causality. The theoretical propositions of procedural justice policing were supported. Future studies are needed with more discernible causal mediation effects.

Key Words: Causal inference; Causal mediation analysis; Police legitimacy; Potential outcome framework; Procedural justice policing; Sensitivity analysis

Introduction

The majority of tests of cause-and-effect relations in the social sciences address the first order question of whether a treatment affects an outcome, and leave unexplored the underlying processes that transmit the putative effect. The failure to focus on mechanisms limits the power and purchase of explanatory frameworks (Bullock, Green, and Ha 2010; Imai et al. 2011). Impact evaluations in criminology, for instance, tend to focus on whether a desired outcome was achieved, not on how that outcome was produced (Famega, Hinkle, and Weisburd 2017). For example, a number of randomised controlled trials (RCTs) have tested the efficacy of hot-spots policing (X), but the lack of assessment of how it transmits its effect (at least partially) through an intervening (mediator) variable (M) to the outcome (Y) means that we do not know how and why hot-spots policing works.

This paper discusses causal mediation analysis as a tool to address this "black-box" view of implementation and causality (Fagan 2017). The contribution of this article is twofold. First, the paper considers the strong assumptions and limitations of the traditional approach to mediation analysis (the product method, see Baron and Kenny 1986) that has been widely used in observational research, especially in the literature of structural equation modelling, where direct and indirect effects are routinely estimated (Mackinnon 2008; Mackinnon, Kisbu-sakarya, and Gottschall 2013). Some users of this method may be unaware of the strong and often unattainable underlying assumptions for estimating indirect effects, which if not met can lead to unreliable and unsound estimates. The current paper demonstrates how to test causal mediation effects using a technique developed by Imai and colleagues to overcome the limitations of traditional approaches to produce potentially causally interpretable results (Imai et al. 2011; Imai, Keele, and Tingley 2010; Imai, Keele, and Yamamoto 2010). The approach includes sensitivity analysis techniques to assess the robustness of results to unmeasured confounding.

Second, this paper uses causal mediation analysis to test a fundamental assumption of the theory of procedural justice policing: namely, that the perceived procedural justice of the police channels the impact of previous contact with the police towards police legitimacy and social identity (for an outline of the models see Figure 1). As a preliminary to that, this paper also shows how to assess the usefulness of – and extract value from – an RCT that experienced a particular form of implementation failure. The Scottish Community Engagement Trial (ScotCET) (MacQueen and

Bradford 2015) was designed to estimate the effect of procedurally just policing on people's experience of procedural justice. But the RCT produced findings contrary to expectations, in that those who received the designed procedurally just treatment reported experiencing lower average levels of procedural justice compared to the control group. Qualitative process evaluations can address what went wrong during implementation (Haberman 2016; MacQueen and Bradford 2017) but such endeavours are retroactive, only focus on startling cases, and can suffer from verification bias. Problematic datasets with unusual results are also often discarded without proper statistical tests having been carried out on the treatment's effects. This paper shows how to test whether value can be extracted by focussing on selection bias, treatment effect inconsistency, and treatment effect heterogeneity – that is, by assessing whether the systematic variation in the dataset is attributable to the research design. To foreshadow the results, an assessment of selection bias, treatment effect consistency, and effect homogeneity supports the idea that the unintended negative treatment effect in ScotCET was produced by the treatment assignment, i.e., that value can be extracted from ScotCET.

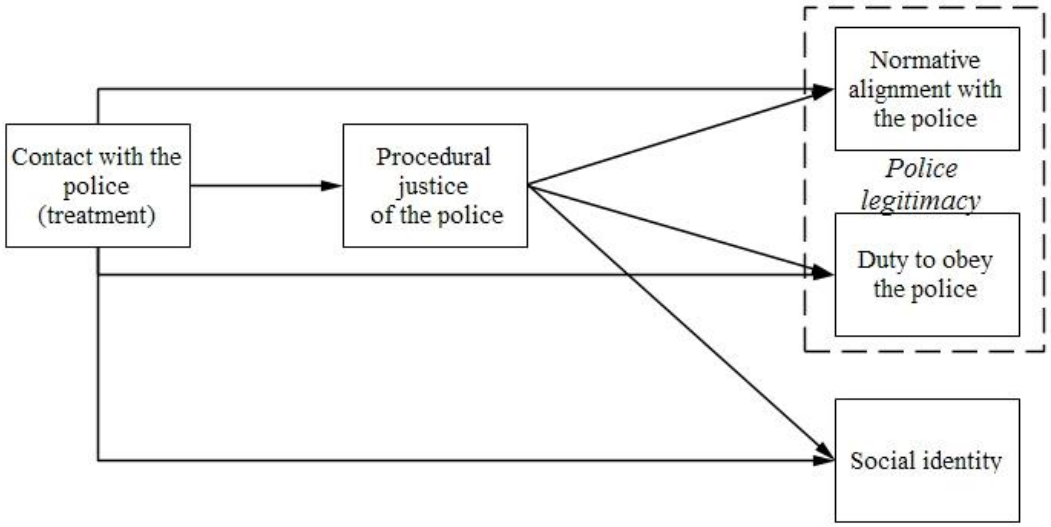


Figure 1: Outline of the tested models

Causal mediation analysis shifts the focus from the total effect of the treatment to the indirect (mediated) effects, hence, experiments with systematic but ambiguous treatments can become interpretable, rendering the initial model testable. Findings from causal mediation analyses support a central prediction of procedural justice

theory, i.e., that the experience of procedural justice mediates the impact of contact with the police on police legitimacy (with moderate levels of robustness to unmeasured confounding) and social identity (with relatively limited robustness to unmeasured confounding).

Procedural Justice Theory and the Scottish Community Engagement Trial (ScotCET)

Procedurally just policing is a topic of much debate in criminology (Tyler, Goff, and MacCoun 2015). Procedural justice theory posits that, when thinking about how the police wield their power and authority, citizens place a good deal of importance on whether officers act – and make decisions – in fair, neutral and respectful ways, and that this process matters more than outcome (Sunshine and Tyler 2003). General perceptions of procedural justice are thought to be influenced by legal socialisation (e.g., Trinkner and Tyler 2016) and direct/vicarious contact with the police (e.g., Bradford 2017; Tyler, Fagan, and Geller 2014). Finally, both the experience and perception of procedural justice are thought to influence people's judgements on the legitimacy of the police as an institution.

Thus far, the evidence base points to the idea that, even in countries as diverse as the US, Australia, Israel, Finland, France, Germany, the UK and China, public concerns about process are more important predictors of police legitimacy than public concerns about effectiveness and fair allocation of outcomes across social groups (Jackson 2018). But as Nagin and Telep (2017) note, the evidence base is dominated by survey-based studies, limiting our ability to estimate causal effects. There have been a few field and laboratory experiments (Murphy and Tyler 2017), and of particular relevance to the current paper is the Queensland Community Engagement Trial (QCET). QCET found that when officers followed a "procedurally fair" script, citizens tended to view their experience with the police as more procedurally just, and that this experience of procedural justice in turn predicted police legitimacy (Mazerolle et al. 2013). ScotCET was designed as a partial replication of QCET and, as QCET, ScotCET tested procedural justice theory in the context of roadside checks, where drivers were stopped by the police for vehicle safety checks and alcohol testing. ScotCET was fielded during the Festive Road Safety Campaign in the December of 2013 and January of 2014 in Scotland, with the design block randomising ten matched pairs of police units to minimise bias across delivery units. Officers in the treatment group were given a series of talking points, with the aim of communicating

procedurally just messages, while officers in the control group carried on with their usual behaviour during these police encounters. After the roadside checks, more than 12,000 questionnaires were handed out to drivers, of which 305 were returned before (122 from the pre-treatment and 183 from the pre-control group), and 510 after the start of the treatment period (176 from the treatment and 334 from the control group). Altogether approximately 6.6% of questionnaires were returned.

In this paper, the analysis links (a) police behaviour in a police-citizen encounter to (b) the subjective experience of procedural justice in that encounter to (c) broader attitudes towards the legitimacy of the police as an institution (Nagin and Telep 2017). Following Hough, Jackson and Bradford (2013; Huq, Jackson, and Trinker 2017), it will be assumed that legitimacy is defined and measured along two connected dimensions. First, normative alignment with the police reflects the degree to which the police respect the societal norms that determine how authority should be rightfully exercised – the inference here is that normative appropriateness justifies the possession of power. Second, duty to obey encapsulates people's willing consent to follow police orders – the inference here is that duty to obey reflects the belief that the police are entitled to make decisions, enforce the law, and dictate appropriate behaviour. A key goal of the current study is to assess the extent to which the putative causal effects of police behaviour on normative alignment and duty to obey are transmitted through the experience of procedural justice. Procedural justice is also posited to mediate the causal effect of police behaviour on citizen social identity. According to procedural justice theory, police officers are representatives not only of the state, but of the communities they serve (Bradford 2014), and if the police treat someone fairly, with respect, and provide citizens with a voice, those citizens will strengthen people's social bonds with that particular community (Murphy and Cherney 2012). Thus, another key goal is to assess to which extent the putative causal effects of police behaviour on social identity is transmitted through the experience of procedural justice.

Before turning to the apparent failure of implementation, I will briefly discuss the measurements used in this paper. There are seven pre-treatment covariates included in all subsequent analyses (unless otherwise noted): age, gender, marital status, educational attainment, employment status, housing, and whether a breath test was conducted by the police during the encounter. Treatment is a binary variable where 0 refers to the control and 1 to the treatment group. Being in the treatment group means

that the respondent had a roadside check with members of the police who were instructed to relay procedurally just messages, whilst in the control group the officers were allowed to carry on with their usual behaviour. All subsequent analyses included this treatment variable, only the data from the treatment period are examined (n=510). Procedural justice, normative alignment, duty to obey, and social identity, were measured using multiple items. They were entered in a confirmatory factor analysis and factors scores were derived for subsequent analysis. For further details regarding the question wording, the confirmatory factor analysis, and the correlation between the different constructs, please refer to Appendix/A. For further information regarding the survey design please consult the appendices of MacQueen and Bradford (2015).

ScotCET's implementation failure

As already mentioned, ScotCET produced the opposite effect to that intended: namely, those who received the treatment reported lower levels of experienced procedural justice compared to the control group (MacQueen and Bradford 2015). In a retroactive qualitative process evaluation designed to find out what happened, MacQueen and Bradford (2017) conducted nine group interviews with police officers who had taken part in the experiment, revealing a number of issues that may have impacted negatively on the treatment implementation. ScotCET coincided with a period of heightened anxiety among officers due to a substantial and unpopular organisational reform in the Scottish police force. Moreover, the participating officers had not been properly briefed regarding the purpose of the study. They had received opaque instructions, assumed that the experiment would have a negative impact on their interactions with members of the public, and felt that the prompts and questionnaire had been assembled by out-of-touch researchers. The focus groups revealed unanimous signs of discontent and negativity towards the experiment. It is conceivable this had a diffuse negative effect on the officers' attitudes and behaviour during encounters in the treatment groups, which may explain (at least partially) the contradictory findings (MacQueen and Bradford 2017).

Despite the problems and apparent failure of implementation mentioned earlier, MacQueen and Bradford (2017) put forward the case that the treatment effect was still interpretable due to the robustness of the study design. In other words, they argue that the treatment and its effects were real, even if both were different in nature from the intentions and expectations of the researchers. This issue is crucial to the

current paper, as it focusses on the extent to which the experience of procedural justice mediates the impact of police-citizen encounters on legitimacy and social identity, thus ScotCET's design and implementation failure needs to be assessed.

MacQueen and Bradford's (2017) claim was mainly based on three considerations. First, there was no selection bias in the original study, where they showed that there was no difference between the control and treatment groups in pre-treatment covariates, either before or during the treatment period (i.e., the randomisation appeared to be successful). Second, the implementation of the treatment did not have an impact on the share of responses in the treatment group (i.e., there was no change in the number of responses received compared to the pre-treatment period, or compared to the control group in the post-treatment period).This would suggest that the overall low response rate of 6.6% does not have an impact on the internal validity of the results, as long as the same kind and proportion of people decide to self-select in the study for both the treatment and control group. Finally, the views regarding the police were on average the same in the control and treatment groups before the treatment period, and they only started to diverge after the treatment implementation (i.e., controlling for all else, the changes can be only attributed to the treatment) (MacQueen and Bradford 2015).

Nonetheless, further research is needed. In particular, police officers reportedly differed in how they had carried out the treatment. Based on their own admissions, some recited the provided messages verbatim, some completely disregarded the prompts, and some only handed out the questionnaires (MacQueen and Bradford 2017). It follows that there are other sources beyond the self-selection bias that might have adversely affected the results. In particular, it is conceivable that (1) the treatment effect varied across the different matched pairs because the officers interpreted and implemented the instructions in different ways (i.e., treatment effect inconsistency) and (2) the treatment had a different impact on certain subgroups, thus leading to biased estimates (i.e., treatment effect heterogeneity). The inherent features of block-randomisation can be harnessed to test both of these potential limitations.

Continuing with the assessment of the apparent implementation failure, to evaluate treatment effect consistency, methods commonly used in meta-analysis can be employed. Each matched pair in ScotCET can be considered an individual study from a meta-analysis. STATA's metaan package was used to perform a random-effect meta-analysis with the treatment as the explanatory variable and all covariates

included in each regression. This random-effect meta-analysis runs the specified regressions for each matched pair, and based on the effect sizes and standard errors, attributes a certain weight to each of them, which is considered when estimating the average treatment effect (ATE). This method also assumes that despite the differences of the underlying effect sizes, all are related through some distribution (i.e., the treatment is specified as a random slope in the model) (Kontopantelis and Reeves 2010). Random-effect meta-analysis also permits the computation of two measures of effect consistency: Cochran's Q and I^2. Cochran's Q is a statistical test for inconsistency, with the null-hypothesis that all studies in the meta-analysis have the same underlying magnitude of effect. Thus, non-significant results indicate consistency of effects. As an additional measure, I^2 is also calculated, which estimates the proportion of variation in the point estimates due to between-study variation. Usually values below 50% are considered as a sign of low inconsistency, while values over 75% considered high (Guyatt et al. 2011; Rhodes, Turner, and Higgins 2016). Due to the lack of control units in one pair, only nine matched pairs were included in the analysis (n=485).

Figure 2 shows a "forest plot" with the treatment's effect on procedural justice across the different matched pairs, and the estimated ATE (also denoted β below) at the bottom (for the forest plots of the other outcomes please refer to Appendix/B). The first three columns of Table 1 summarises the results from the analysis. The treatment has a significant negative effect on procedural justice (β=-0.435, $p<0.05$) and duty to obey (β=-0.579, $p<0.05$), however the rest of the effects are not significant. These are in line with the findings of the original study (Macqueen and Bradford 2015), and suggest that the contact with the officers in the treatment group diminished people's views about the police compared to the encounters in the control group. Importantly, Cochran's Qs are not significant, and the I^2s show either low (duty to obey: I^2=43.06%; social identity: I^2=42.42%) or minimal (procedural justice: I^2=2.1%, normative alignment: I^2=8.8%) inconsistency. This lack of inconsistency across delivery units implies that even if the police officers acted in a different manner, the impact of their interactions during the police stops was fairly similar across the locations.

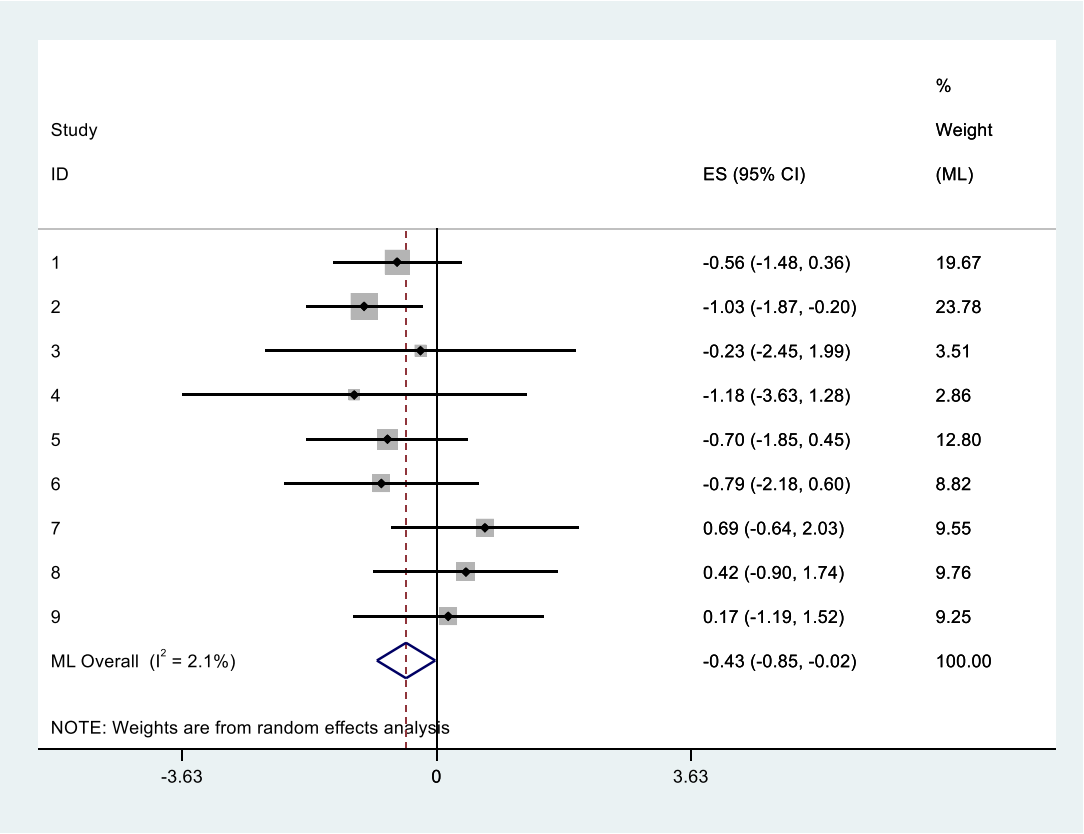


Figure 2: Treatment effect consistency for procedural justice

	ATE	*Cochran's Q*	I^2	*Covariate heterogeneity differences*	*Design heterogeneity differences*	*Treatment-covariate interaction*
Procedural justice	-0.435* [-0.852, -0.018]	7.99	2.1%	0.016	0.006	NS
Normative alignment	-0.257 [-0.646, 0.133]	10.1	8.8%	0.035	0.015	NS
Duty to obey	-0.579* [-1.128, -0.030]	15.18	43.06%	0.038	0.009	NS
Social identity	-0.262 [-0.558, 0.034]	15.26	42.42%	0.033	0.007	NS

**p<0.05, **p<0.01*

Table 1: Average treatment effects from the random-effects meta-regression, Cochran's Q, I^2, design and covariate heterogeneity, and treatment-covariate interactions (NS = not significant)

A second potential complication in this study is the treatment's systematic variation across subgroups within the population. In case of such heterogeneity, the assumption that the ATE is same for each individual might not be tenable, and thus the various estimators of the treatment effect might be altered even in the absence of selection or confounding bias (Na, Loughran, and Paternoster 2015). The block-randomised design permits two different analyses of effect heterogeneity: (1) treatment effect heterogeneity, which scrutinises the ATE's dependency on pre-treatment covariates and (2) design heterogeneity, where in addition to the pre-treatment covariates, the treatment's dependence on the different blocks is also testable. Because there was no initial expectation with regards to treatment-covariate and treatment-matched pair interactions, an automated solution, the "FindIt" R package and Squared Loss Support Vector Machine (L2-SVM) (Imai and Ratkovic 2013) was applied. This L2-SVM model first rescales the covariates (using a LASSO-regularisation), then fits the model (again, with a series of iterated LASSO fits) by also relying on generalised cross-validation statistics. This approach automatically tests the potential interactions between the various covariates in the model, as well as the interaction between the covariates and the treatment, only flagging the influential ones. Two L2-SVM models were fitted for each outcome and were subsequently compared to each other and to the ATE. The first model only considered the covariates, the second one both the covariates and the blocking design. As indicated by the fourth column in Table 1, accounting for the treatment effect heterogeneity only lead to limited changes in the ATE, with alterations in the point estimates ranging from 0.016 to 0.038. The fifth column shows that after adding the matched pairs to the analysis, these differences drop even further, with miniscule changes ranging from 0.006 to 0.015. This drop is anticipated, since the blocking was designed to account for the sampling variability. Finally, no treatment-covariate interaction emerged in either of the models. The lack of interactions and the small changes in the ATEs indicate that the treatment effect can be considered by and large homogeneous. This means that the treatment's effect from ScotCET had very similar impact in the population, and that there were no subgroups which were more or less receptive to its influence, or delivery units that had disparate impact on the results.

To summarise, the examination of selection bias, treatment effect inconsistency and treatment effect heterogeneity provided strong evidence regarding the internal validity of the treatment's effect. These demonstrated that the same kind of people

answered the surveys in the treatment and control group, that they were affected in a very similar way by the treatment across the matched pairs, and that the treatment's effect did not vary across the subgroups either. Thus, these tests all substantiate MacQueen and Bradford's (2017) assertion about the robustness of the research design.

The preliminary analysis conducted so far suggests that the treatment has a real, consistent, and homogeneous effect on procedural justice and various other outcomes. Nevertheless, it is still very difficult to give a proper definition for the treatment which produced such unintended effects. Therefore, this article proposes a mere descriptive interpretation, assuming that the police encounters were significantly different in the treatment and control group, hence interpreting the treatment as a systematically different contact with the police. For the treatment group, the experience during the encounter with the police was on average more negative compared to the control group, most likely due to meeting disgruntled officers. With all these considered, this paper proceeds to examine the mediating effects and tests a fundamental question found in the procedural justice literature: whether the impact of a person's previous positive/negative contact with the police is channelled through procedural justice to affect certain outcome variables (e.g., legitimacy).

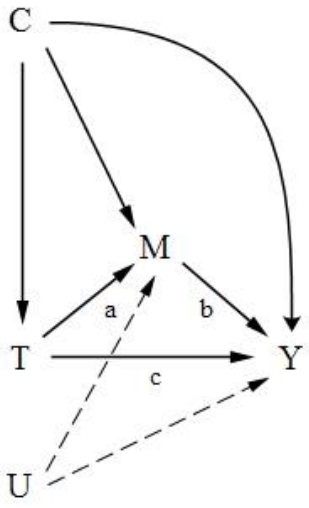


Figure 3: Outline of a mediation model with a single mediator

Causal mediation analysis

Classical definitions of direct and indirect effects

In this article I hypothesise that the quality of contact with the police (X) shapes respondents' attitudes (regarding procedural justice) (M), which in turn influences – among other things – their views on the legitimacy of the police (Y). Because traditionally X refers to any kind of (even observed) variable, this paper will denote the antecedent variable as T, which indicates the randomised treatment. In addition, it is conventional to control for a vector of pre-treatment covariates C (see Figure 3). Using the traditional decomposition of the product method, and as depicted by Figure 3, 'c' is a regression coefficient that stands for the direct effect of T on Y, while the product of 'a' and 'b' (i.e., the estimates of T's effect on M, and M's effect on Y) stands for the indirect effect of T that goes through M towards Y. This approach is generally referred to as the product method as an indication of how the indirect effect is derived. Following Baron and Kenny's (1986) seminal article, product method mediation analysis with a single mediator can be expressed as:

$$(1) \qquad M=\beta_0+\beta_1 t+\beta_2 c+\varepsilon_1$$
$$Y=\theta_0+\theta_1 t+\theta_2 m+\theta_3 c+\varepsilon_2$$

In the first equation, β_1 denotes the effect of the treatment on the mediator ('a' in Figure 3) after taking into account the covariates (β_2) with the intercept (β_0) and error term (ε_1) In the second equation, θ_1 is the direct effect of T on Y ('c' in Figure 3) after controlling for M (θ_2) ('b' in Figure 3) and C (θ_3) with the constant (θ_0) and error terms (ε_2). The mediated (indirect) effect is the product of the coefficient of the treatment in the regression for the mediator (β_1) and the coefficient of the mediator in the regression for the outcome (θ_2).

However, several criticisms have emerged regarding the applicability of the product method. First, the product method is only capable of identifying[1] direct and indirect effects if the linearity assumption holds (Imai et al., 2010b; Jo 2008). This

[1] Identifiability here – and throughout the paper – means that an (causal mediation) effect is consistently estimable. It follows that identification is a necessary, but not sufficient requirement, which precedes the actual statistical estimation and refers to the ability to obtain the effects of interest (Manski 2007; Keele 2015). Importantly, this is different from the model-based identification regularly used in the structural equation literature.

means that for non-linear (e.g., multinomial) models the indirect effect cannot be computed relying on the product method. The second caveat is usually referred to as no-interaction assumption. This prescribes that there cannot be an interaction between the treatment and the mediator which affects the outcome. The absence of interaction is important, because it permits the effect decomposition and also provides a good indication for effect homogeneity, which is a further requirement (i.e., the causal effects are constant across cases) (Kline 2015). In the presence of an interaction (e.g., between the treatment and procedural justice in this paper), the method of identification of the direct and indirect effects breaks down as it becomes unclear how to calculate the total effect. Yet, the lack of interaction is not sufficient, because effect homogeneity needs to apply to each individual case, which is an untestable (and highly unlikely) assumption.

A further limitation concerns the applied literature rather than the method itself. Similarly to other causal techniques, causal mediation analysis relies on no unmeasured confounder assumptions which are usually addressed by the random assignment of participants to treatment and control group(s). In other words, if we randomly assign people to a treatment or control group, we can safely assume that they will not differ across important and influential measured *and* unmeasured characteristics (e.g., age, education, previous experience with the police), and hence the exogeneity assumption is met. However, even if the treatment T is randomly assigned, the mediator-outcome relationship is not randomised, which might result in people self-selecting for their mediators independent of the treatment and due to an unmeasured confounder U (depicted in Figure 3). This U can generate biased direct and indirect effects thus producing unreliable results. This issue has been mostly overlooked, partly because it was not discussed in the classic article by Baron and Kenny (1986); although it was discussed in an earlier paper of one of the authors, Judd and Kenny 1981).

To further complicate matters, randomisation of the mediator, as proposed by some (Bullock et al. 2010; Spencer et al. 2005; Walters and Mandracchia 2017), is not sufficient either for assessing the indirect effect. When both the mediator and treatment are randomly assigned, the exogeneity assumption is satisfied for each, however, it does not apply to the combination of the two. In such cases, the treatment can causally affect the mediator, and the mediator can causally affect the outcome, however, the mediator does not transmit the effect of the treatment anymore due to its random

assignment (Imai, Keele, and Tingley 2010; Keele 2015). Thus, this is a germane problem in the literature as special design-based strategies need to be applied for a better chance of identifying causally mediated effects (Imai et al. 2011; Imai, Tingley, and Yamamoto 2013; Pirlott and Mackinnon 2016). A careful selection of pre-treatment covariates might mitigate the possibility of an unmeasured influential U, but it can rarely solve the issue altogether (VanderWeele 2015). To better understand the assumptions and estimation needed for causal mediation, it is crucial to introduce a more general definition of direct and indirect effects.

Counterfactual definitions of the direct and indirect effects

In the following paragraphs the controlled direct effect, natural direct effect, and natural indirect effect are discussed as more general definitions of the direct and indirect effects from the product method. These new, general definitions rely on the potential outcome framework and counterfactual way of thinking (Pearl 2001; Robins and Greenland 1992).

These counterfactual definitions are given assuming a binary treatment variable T, mirroring the one used in ScotCET. For all of these counterfactual definitions, let us assume that we compare two hypothetical worlds where in the first world T is set to 0 (i.e., control) and in the second T is set to 1 (i.e., treatment) within the same individual at the same moment in time. Using ScotCET as an example, this would mean that the same person would have been exposed to both the treatment and the usual police practice during the roadside check at the very same moment in time from the very same officer(s). Although in real life we can never know what would have happened to that individual had that person been assigned to the other group[2] instead of the observed one, hypothetically we can conceive these two separate counterfactual outcomes. It follows that counterfactual inference can never be derived for a single individual, only for a population. Thus, for the general definitions of the effects, the language of conditional expectations is employed, to indicate that population average effects are, in fact, conditional expectations of the individual level effects. For example, in the analysis of ScotCET we may take this population to be the respondents

[2] This limitation is often referred to as the fundamental problem of causal inference (Holland 1986).

of the study for whom the variables are observed. All the expected values E(.) of random variables below denote expectations over distributions in this population.

There are different commonly used ways of defining direct and indirect effects in this framework. They differ mainly in what values are considered for the mediator variable M. The controlled direct effect (CDE) considers a specified value of M=m and defines the direct effects as:

$$CDE(m)=E[Y(1,m)-Y(0,m)] \quad (2)$$

This thus captures the expected increase in Y when T changes from T=0 to T=1 while M is held at the value m for everyone (i.e., within the individual M is kept constant, while she receives both the control and treatment at the same time). This is a direct effect since the effect of T is not transmitted through M. The value of CDE might change depending on the chosen value of m, which also means that relying on CDE does not allow the decomposition of the total effect to direct and indirect effects. Still, setting the m to different values can provide policy relevant information, such as the number of meetings people on parole should attend in order to reduce their recidivism.

The natural direct effect (NDE) is defined as

$$NDE=E[Y(1,M(0))-Y(0,M(0))] \quad (3)$$

This is similar to the controlled direct effect, in that it estimates the expected increase in Y when T changes from T=0 to T=1. However, the NDE does not hold the value of M constant, instead it permits it to take its value in the "natural" way for each individual if that individual had been assigned to the control condition. This modification allows for the decomposition of the effects.

The natural indirect effect (NIE) is defined as:

$$NIE=E[Y(1,M(1))-Y(1,M(0))] \quad (4)$$

It does the opposite of NDE as it approximates the expected increase in Y when the treatment is kept at T=1, while M is freed to take its natural value for the treatment and the control group respectively. This is an indirect effect that captures the effect of T on Y which is transmitted through M:

Importantly, both the direct and indirect effect can be defined through holding M at its potential outcome given T=1 for the direct effect, while holding Y at its potential outcome T=0 for the indirect effect:

(5) $NDE_{alt}=E[Y(1,M(1))-Y(0,M(1))]$

(6) $NIE_{alt}=E[Y(0,M(1))-Y(0,M(0))]$

This will produce identical results in respect of the total effect, as shown in (7) below. However, these alternative definitions differ in where the effect of the potential T-M interaction term is assigned (Daniel et al. 2015; Muthen and Asparouhov 2015). Using the classic definition of NIE and NDE in (2)-(3) (Pearl 2001), the interaction term is assigned to the indirect effect, while for the NDE_{alt} and NIE_{alt} in (5)-(6) it is assigned to the direct effect. To avoid confusion, sometimes the words "total" and "pure" are added to the direct and indirect effects, where total indicates the added interaction effect. Therefore, the NIE is the total indirect effect (TNIE), while the NDE is the pure direct effect (PNDE). Conversely, the alternative definitions of NIE_{alt} and NDE_{alt} refer to the pure indirect (PNIE) and total direct effects (TNDE) respectively.

Using either of these definitions of natural effects, the total effect (TE) of T on Y can be decomposed as the sum of direct and indirect effects, i.e.:

(7)
$$\begin{aligned}
&TE=E[Y(1)-Y(0)]=\\
&E[Y(1,M(1))-Y(0,M(0))]=\\
&\{E[Y(1,M(1))-Y(1,M(0))]\}+\\
&\{E[Y(1,M(0))-Y(0,M(0))]\}=\\
&NIE+NDE=NIE_{alt}+NDE_{alt}=\\
&TNIE+PNDE=PNIE+TNDE
\end{aligned}$$

As described above, the identification of the direct and indirect effects through the potential outcome framework does not posit the no-interaction assumption, which allows for the effect decomposition even in the presence of such an association. Moreover, it is nonparametrically identifiable, thus does not require the linearity assumption either, which permits more flexible modelling (Pearl 2001).

Estimation of the natural direct and indirect effects

To estimate the kinds of effects defined above, we first specify models for Y given T, M, and C and for M given T and C, estimate these models using the observed data, and apply formulas which are analogous to (2)-(6) to these fitted models. This produces estimates of the direct and indirect effects, if certain assumptions are satisfied (these assumptions are discussed in the next section).

To illustrate this idea, suppose that we consider the models given in equation (1), but now with the added interaction between T and M, θ_4, assuming the linearity of the effects. Notice that unlike in (1) the error terms are no longer present as they are expected to be $E(\varepsilon)=0$ in the equations. Provided certain assumptions hold for the respective effects, on average for the population, the following can be derived:

(8) $CDE(t_1, t_0;m)=(\theta_1+\theta_4 m)(t_1-t_0)$

(9) $NDE(t_1, t_0; t_0)=(\theta_1+\theta_4(\beta_0+\beta_1 t+\beta_2 c))(t_1-t_0)$

(10) $NIE(t_1, t_0; t_1)=(\theta_2\beta_1+\theta_4\beta_1 t)(t_1-t_0)$

where 't_0' and 't_1' denote the values of T in the treatment and control groups respectively. These are estimated by substituting estimated values of the parameters on the right-hand side of (8)-(10). This also shows how the extended definitions can easily accommodate models for Y with interactions between T and M. From these formulas it can be easily discerned that when $\theta_4=0$, (8) and (9) coincide ($CDE(t_0,t_1;m)=NDE(t_0,t_1;t_1)=\theta_1(t_0-t_1)$), and (10) is simplified to the traditional product method ($NIE(t_0,t_1;t_0)=\theta_2\beta_1(t_0-t_1)$). It follows that the product method is a special case of causal mediation analysis which is obtained under assumed linear models with no interaction (Imai et al. 2011).

As an alternative to these fully parametric models Imai et al. (Imai et al. 2011) have proposed a semiparametric estimation approach. Following their modelling strategy firstly, two regression models are fitted for the mediator and the outcome of interest, similarly to the parametric approach. Likewise, two sets of mediator (conditional on T and C) and outcome (conditional on M, T, and C) values are generated for every observation for each level of treatment $T=t_0$ and $T=t_1$. Again, in a similar vein, the effects are computed through averaging the differences between the predicted potential values. This approach is superior to the previous one in that it is applicable for any kind of link function, while the parametric one is only applicable to

a couple of special link functions (i.e., linear and binary logit with rare outcome variables) (VanderWeele 2015).

Because of its flexibility, here the semiparametric approach was used but, notably, for linear outcome variables, the two approaches will generate almost identical results. Finally, both approaches recommend using resampling techniques, such as the nonparametric bootstrap or Monte Carlo approximation to correctly represent the prediction uncertainty of the estimates in these models. For all models in this paper, the treatment was binary, and the mediator and outcome variables were continuous, with all covariates included in the models. The "mediation" R package (Tingley et al. 2014) was used with interaction allowed between the treatment and the mediator, and 1000 bootstrap replicates were specified for estimation of standard errors.

Assumptions of causal mediation analysis

In order to make causal claims using the estimators outlined above the sequential ignorability assumption needs to be satisfied (Imai et al. 2010a). This no unmeasured confounder assumption lists the different sources of unmeasured confounders U that can produce biased results and requires that, after controlling for all pre-treatment covariates C, there is no unmeasured confounder for:

a) The relationship between the treatment (T) and outcome (Y)
b) The relationship between the mediator (M) and outcome (Y)
c) The relationship between the treatment (T) and mediator (M)
   and,
d) There is no post-treatment mediator-outcome confounder (L) that was affected by the treatment

From these four assumptions, (a) and (c) constitute exogeneity assumptions usually applied to determine the average treatment effect in randomised experiments, and are automatically satisfied in the case of random assignment of T (as it was done with ScotCET). For (b) to be fulfilled M either needs to be as-if-randomly assigned (using data from special research designs which are not considered here) or assumed that it is as-if randomly assigned after controlling for T and C. To accomplish the final point (d) one needs to rely on a parsimonious model similar to Figure 3, as it posits

that there cannot be other post-treatment confounders (essentially other mediators) that are not included in the model. In terms of the new definitions of the different direct and indirect effects assumptions, (a) and (b) are sufficient to derive the CDE(m)[3], while (a)-(d) are needed for the NDE and NIE. Finally, as with randomised experiments in general, the stable treatment unit value assumption also needs to be met.

Sensitivity analysis

Similarly to other techniques in the causal inference literature, causal mediation analysis also relies on untestable and non-refutable assumptions (Manski 2007). Although, the strong claims of the sequential ignorability assumption cannot be directly tested on the observed data, sensitivity analyses can be utilised that permit researchers to quantify the robustness of their findings and assess the potential influence of unmeasured confounders. Critically, even if the treatment was randomised, the ignorability of mediator M should be studied through evaluating whether there is a reasonable chance that omitted variable U might invalidate the results (see assumption b above). However, in most cases sensitivity analyses will not provide easily discernible results, rather a range of values that will indicate the plausibility of the results. As there are no established benchmarks upon which one could decide on the absolute robustness of results, inferences must be informed by previous findings from the field and should be compared with the impact of other measured confounders. There are several different sensitivity analysis techniques (Ding and Vanderweele 2016), of which two will be discussed here. These techniques work especially well with continuous mediators and are capable of gauging the robustness of the NDE and NIE.

The first technique (Imai et al. 2010a; Imai et al. 2011; Imai and Yamamoto 2013) fits two regressions, one for M and the other for Y with a T-M interaction. One can take the error terms (ε) from these regressions and specify a correlation between them denoted by ρ. Since the error terms incorporate the impact of U, the value of ρ will relatively increase if there is an influential U that affects both M and Y. Conversely, ρ will comparatively decrease in the absence of an influential U. Thus, the

[3] Notably, the usual regression-based models will no longer be sufficient, other approaches, such as marginal structural models, structural nested models and so on, can be used to derive the CDE (Coffman and Zhong 2012; Lepage et al. 2016; Moerkerke et al. 2015).

sensitivity of the mediation results can be tested by systematically increasing the correlation between the two εs and evaluating the extent to which the estimates are altered. Accordingly, the direct and indirect effects will be the functions of the parameter ρ, and the higher value ρ takes will imply relatively more robust results. A mathematically equivalent but perhaps more intuitive way of reporting the results is to consider the R-squared statistics and interpret the results in terms of U's explanatory power. There are two R^2s worthy of interest. The R^2 for the residual variance shows the proportion of previously unexplained variance that is explained by U. Alternatively, the R^2 for the total variance represents the same, but for the proportion of the original variance. In the case of the R^2s, higher values will indicate relatively lower sensitivity to the violation of the sequential ignorability assumption compared to results from similar studies.

The other sensitivity analysis technique is called the left out variable error method (LOVE) (Cox et al. 2013; Mackinnon and Pirlott 2016), which assesses the extent to which an unmeasured variable U would have to affect the association between M and Y in order for the observed association to be attributable to this confounding alone. This approach classifies the error due to U as a misspecification error and applies correlation techniques for bias detection. Therefore, LOVE relies on the correlation between T-M, T-Y, and M-Y to approximate the correlation between U-Y and U-M. The average of the U-Y and U-M correlation corresponds to a correlation coefficient that would make the observed mediated effect zero. As in the previous case, a higher coefficient will entail less sensitive results. The major advantage of this method is that it enables a less convoluted assessment of the effect of U on the M-Y relationship. However, this straightforwardness comes at price: unlike the previous sensitivity analysis, the LOVE technique does not include pre-treatment covariate Cs, which considerably limits its authenticity for the model under scrutiny. Nevertheless, the LOVE method can be still a powerful detector of bias and an easy check of the relationships between T, M, and Y.

Results

The causal mediation analysis results are displayed in Table 2. For each model, the treatment (T) is a binary variable representing the encounter with the officer(s) from the treatment or control group, the mediator (M) is procedural justice, and the outcome (Y) is either normative alignment with the police, duty to obey the police, or social

Procedural justice as mediator	*Type*	*Average effect*	*Mediate %*	*Mean ρ*	*Residual R^2*	*Total R^2*	*Mean LOVE*
Normative alignment	NIE_{mean}	-0.207* [-0.384, -0.031]	84.2%	0.6	0.36	0.20	0.7
	NDE_{mean}	-0.007 [-0.261, 0.240]		~0.1	0.01	~0.01	
Duty to obey	NIE_{mean}	-0.153* [-0.297, -0.018]	34.9%	0.5	0.25	0.17	0.7
	NDE_{mean}	-0.279* [-0.540, -0.008]		0.7	0.49	0.32	
Social identity	NIE_{mean}	-0.052* [-0.108, -0.005]	16.9%	0.3	0.09	0.12	0.5
	NDE_{mean}	-0.243* [-0.411, -0.080]		0.8	0.64	0.46	

**p<0.05, **p<0.01*

Table 2: Causal mediation analysis results with averaged NDE and NIE effects and sensitivity analyses

identity. Both the natural direct effect (NDE) and natural indirect effect (NIE) in Table 2 take the average of the direct and indirect effects estimated in (3) and (5) and (4) and (6) respectively. I use the model fitted for normative alignment (first row) to exemplify the interpretation of the results. The NIE is -0.207, which is significant on the 5% level. This NIE shows that procedural justice mediates 84.2% of the total effect with a non-significant natural direct effect of -0.007. To nullify the NIE the mean correlational coefficient between the error terms from the model for the mediator and outcome would need to be 0.6. This (ρ=0.6) corresponds to 36% of the residual variance and 20% of the total variance of the model. Thus, this relationship seems to be less sensitive or, in other words, fairly robust to unmeasured confounding. By contrast, for the NDE's effect to reach zero, this correlation coefficient would only need to approach 0.1, with the power to explain 1% of the residual variation and less than 1% of the total variation. Therefore, this result is highly sensitive to unmeasured confounding, which

corresponds to its NDE value that is close to zero and non-significant. Finally, the left-out-variable value (LOVE) implies that on average an unmeasured confounder would need to have a 0.7 correlation with the mediator and outcome to make the NIE non-significant.

Procedural justice seems to channel the effect of the treatment to normative alignment (as discussed in the previous paragraph), duty to obey (NIE_{mean}=-0.153, $p<0.05$, Mediate %=34.9%, $\rho=0.5$, $R^2_{residual}=0.25$, $R^2_{total}=0.17$, LOVE=0.7) and social identity (NIE_{mean}=-0.052, $p<0.05$, Mediate %=16.9%, $\rho=0.3$, $R^2_{residual}=0.09$, $R^2_{total}=0.12$, LOVE=0.5). In case of normative alignment, the treatment does not have a significant direct effect, whilst for both duty to obey (NDE_{mean}=-0.279, $p<0.05$, $\rho=0.7$, $R^2_{residual}=0.49$, $R^2_{total}=0.32$) and social identity (NDE_{mean}=-0.243, $p<0.05$, $\rho=0.8$, $R^2_{residual}=0.64$, $R^2_{total}=0.466$) the direct effect is not only significant, but stronger than the indirect effect.

Notably, and despite the difference in the magnitude of the effect size of the NIE, normative alignment and duty to obey both have the same LOVE-score and very close ρ scores for their NIEs, indicating similar levels of robustness to unmeasured confounding. In comparison, social identity's NIE appears to be more sensitive.

Finally, the inclusion of the interaction effect needs to be discussed. Another improvement of causal mediation analysis is that it manages to resolve the inclusion of the interaction effect while still guaranteeing a meaningful decomposition. In Table 2 the average NIE and NDE were included. By contrast, Table 3 has the NIEs and NDEs discussed in the methodological overview: NIE corresponds to (4), NDE to (3), while NIE_{alt} corresponds to (6), and NDE_{alt} to (5)[4].Taking normative alignment as an example, when the whole interaction is attributed to the indirect effect (NIE), it has an effect size of -0.244, mediating almost fully the effect of the treatment (Mediate %=98.9%), with a $\rho=0.7$ needed to make the indirect effect non-significant, with 49% of the residual, and 25% of the total variation explained. Conversely, if none of the interaction is attributed to the mediated effect (NIE_{alt}), it has an effect size of -0.171, procedural justice only mediates a little more than two-thirds of the treatment's effect

[4] As noted earlier, the different decompositions will refer to the same total effect. For instance, for normative alignment it will be: TE=-0.215=NDE_{mean}+NIE_{mean}=-0.007+-0.207=NIE+NDE=-0.244+0.029= NIE_{alt}+NDE_{alt}=-0.171+-0.044.

(Mediate %=69.5%), with a mean ρ=0.5, which coincides with the residual variance of 25%, and the total variance of 13%.

Procedural justice as mediator	*Type*	*Effect size*	*Mediate %*	*Mean ρ*	*Residual R^2*	*Total R^2*
Normative alignment	NIE_{alt}	-0.171* [-0.321, -0.026]	69.5%	0.5	0.25	0.13
	NIE	-0.244* [-0.449, -0.037]	98.9%	0.7	0.49	0.25
	NDE	0.029 [-0.231, 0.284]		0.1	0.01	0.01
	NDE_{alt}	-0.044 [-0.299, 0.213]		0.2	0.04	0.02
Duty to obey	NIE_{alt}	-0.130* [-0.260, -0.014]	29.7%	0.4	0.16	0.11
	NIE	-0.176* [-0.345, -0.020]	40.2%	0.5	0.25	0.16
	NDE	-0.256 [-0.514, 0.009]		0.7	0.49	0.32
	NDE_{alt}	-0.302* [-0.558, -0.031]		0.7	0.49	0.32
Social identity	NIE_{alt}	-0.029 [-0.074, 0.001]	9.2%	0.1	0.01	0.01
	NIE	-0.075* [-0.156, -0.006]	24.7%	0.4	0.16	0.11
	NDE	-0.219* [-0.387, -0.054]		0.8	0.64	0.46
	NDE_{alt}	-0.295** [-0.472, -0.124]		0.8	0.64	0.46

**p<0.05, **p<0.01*

Table 3: Causal mediation analysis results with the interaction's effect attributed either to the NIE or NDE, and sensitivity analyses

Even if it is difficult to determine where to assign the effect of the interaction, Table 3 can help to inform the researcher about the presence/absence of an influential T-M interaction. Simply based on the magnitude of change in the effect size, moral alignment is the most affected by the allocation of the interaction. However, in case of duty to obey a smaller change influences the significance of the NDE. Similarly, in case of social identity the significance of the NIE is dependent on the assignment of the interaction effect. These examples underline the importance of including the interaction in the analysis, and the limitations of the product method which would not have accounted for the impact of the T-M interaction.

Discussion

Much empirical research in the social sciences is focussed on identifying causal relationships, and this is especially true for experimental studies. Yet, most of these efforts only scrutinise the average causal effects, they are not concerned with underlying causal processes and mechanisms. This article has discussed causal mediation analysis as a promising statistical method to "pry open" this black box of causality. This approach goes beyond the traditional product method and can be applied to models with non-linear link functions and interactions, without positing the effect homogeneity assumption, while quantifying the potential influence of unmeasured confounders for the mediator-outcome relationship through sensitivity analyses (Imai et al. 2011; Imai, Keele, and Tingley 2010; Imai, Keele, and Yamamoto 2010). Unlike in previous criminological work, where causal mediation analysis has been used in a longitudinal research context (Walters 2015, 2017), here it is employed in an experimental setting. Moreover, this paper went beyond a recent review of applied literature on causal mediation in criminology (Walters and Mandracchia 2017) by (a) presenting a versatile statistical technique and (b) utilising the potential outcome framework to outline fundamental causal assumptions and describe new definitions of direct and indirect effects. Furthermore, it recommends two sensitivity analysis methods that can be easily used in most applied settings.

To exemplify the utility of causal mediation analysis, this paper chose to reanalyse the ScotCET dataset. The assessment of the selection bias, treatment effect consistency, and effect homogeneity shows that the treatment effect does not affect people's self-selection in the study, that it is very similar across the matched pairs, and that there is small covariate and minimal design heterogeneity. Thus, and despite the

apparent failure of implementation, these results indicate that the effects of the treatment on various outcomes is identifiable and was produced by the experimental design. It follows that the results emerging from ScotCET can be harnessed for further analysis. Conducting similar tests of selection bias, treatment effect inconsistency, and heterogeneity, for other experiments with block-randomised designs should be common practice and imperative in examining the identifiability of the ATEs.

The potential outcome framework used in this article is a rigorous tool making modelling assumptions explicit and offering new definitions of direct and indirect effects, which can be identified based on whether particular assumptions are satisfied. Future research would benefit from considering each step of the sequential ignorability assumption, and gauging whether the proposed causal mediation models are identifiable. Sensitivity analysis techniques would provide further insight into the robustness of emerging results, and could make tenuous relationships easily affected by third common causes (Nagin and Telep 2017) more discernible. At times, when parts of the experimental community is preoccupied with the "replication crisis" and "p-hacking", these sensitivity analysis techniques could be readily applied as further tests regarding the viability of results.

Causal mediation analysis allows a change in the focus of the analysis, moving from the treatment effect to the mediated effect of procedural justice of the police. The rich set of pre-treatment covariates from the ScotCET dataset allowed a robust test of the theory of procedural justice policing. Procedural justice appears to channel the impact of contact with the police towards moral alignment, duty to obey, and social identity, although at different levels of sensitivity. It is notable that in the case of duty to obey and social identity, the direct effect remained significant, indicating that not all aspects of the contact's impact are mediated by procedural justice.

As with every method, causal mediation analysis faces certain challenges that need to be addressed. Even with a randomised treatment, the sequential ignorability assumptions are very demanding. For instance, in case of ScotCET, there might be influential covariates that were not measured and thus not included in the models (e.g., earlier contact with the police, victimisation). Unlike in other fields, such as epidemiology, where dozens of pre-treatment covariates are regularly considered, in the social sciences it is usually very difficult to find exhaustive lists of such covariates (VanderWeele 2015). Moreover, the results of the sensitivity analyses cannot be assessed on their own, but only with regard to the list of pre-treatment covariates that

are accounted for. Noticeably, some of the results become more robust to unmeasured confounding when the covariates are not included in the models (see: Appendix/C Table 3/a). This means that the robustness of the results can only be determined in comparison to other variables in the models, unless sensitivity benchmarks have been established.

Another potential criticism of causal mediation analysis is that it requires the assumption that only a single mediator will channel a treatment's effects towards the outcome. Yet, in the social sciences, theories often posit multiple pathways. In non-Western countries, for example, police effectiveness is usually considered alongside procedural justice (Bradford et al. 2014). However, this would violate assumption (d) of the sequential ignorability assumption, which does not allow the presence of further mediators. Hence the method presented here can only be applied to relatively simple models, and other more complex solutions need to be pursued when multiple mediators are present (Daniel et al. 2015; VanderWeele and Vansteelandt 2014).

Finally, this study's treatment merits some discussion. Even though the diagnostics of selection bias, treatment effect consistency and homogeneity indicate that the treatment's effect is only attributable to the design, still without knowing exactly what transpired during the roadside encounters, only a descriptive interpretation can be provided, which renders any explanation of the direct effects ambiguous. Moreover, it is plausible that the treatment effect without the discussed implementation failure would have produced different results. As with other experimental results, multiple trials are needed to revisit the findings presented here. Yet, by relegating the treatment's effects and elevating the mediated effects, causal mediation analysis permitted a clarification regarding to what extent these experiences were mediated by procedural justice, thus producing theoretically valuable findings.

Appendix/A – Measurement

In this paper, several different constructs were measured with multiple items: procedural justice (4 items), normative alignment (3 items), free duty to obey (3 items), and social identity (2 items). The question wording and response alternatives are all detailed in Table 1/a.

Construct	*Items*	*Response alternatives*
Procedural justice	The police in Scotland make fair decisions. The police in Scotland listen to people before making decisions. The police in Scotland treat people with dignity and respect. The police in Scotland treat everyone equally.	1 – Hardly ever 2 – Not very often 3 – Some of the time 4 – Most of the time
Normative alignment	The police have the same sense of right and wrong as me. The police stand up for values that are important for people like me. I support the way the police usually act.	1 – Strongly disagree. 2 – Disagree. 3 – Neither agree nor disagree 4 – Agree 5 – Strongly agree
Duty to obey	I feel a moral obligation to obey the police. I feel a moral duty to support the decisions of police officers, even if I disagree with them. I feel a moral duty to obey the instructions of police officers, even when I do not agree with them.	
Social identity	I see myself as a member of the Scottish community.	

	It is important to me that others see me as a member of the Scottish community.

Table 1/a List of constructs, measures, and response alternatives

All constructs with multiple items were entered in a confirmatory factor analysis, the results are depicted by Figure 1/a. According to the model fit indices (CFI=0.977, TLI=0.968, RMSEA=0.056, SRMR=0.033) the model fit the data well. The factor loadings were relatively high (λ=0.629-0.916) for all latent variables which implies that the measurement models performed well. After the confirmatory factor analysis, factor scores were derived and used in all subsequent analysis.

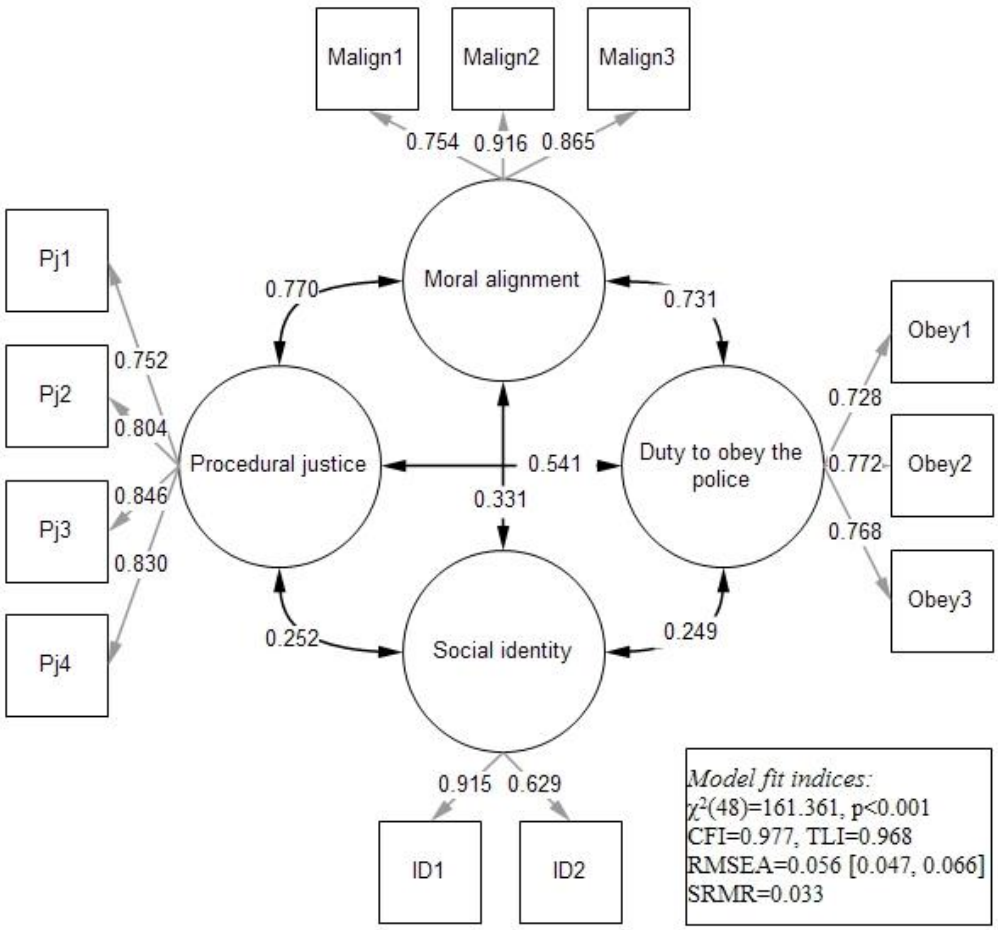


Figure 1/a Confirmatory Factor Analysis of the Constructs Used in the Article (All relationships are significant on the p<0.001)

Correlational results

The correlational results (Table 2/a) show that the treatment had a weak negative association with the other variables. The correlation between treatment and social identity emerged with the biggest magnitude ($r=-0.150$, $p<0.05$), followed by duty to obey ($r=-0.144$, $p<0.01$), normative alignment ($r=-0.114$, $p<0.05$), and procedural justice ($r=-0.103$, $p<0.05$).

The mediator of interest, procedural justice, followed the expected pattern: it had a strong positive correlation with normative alignment ($r=0.698$, $p<0.01$) and duty to obey ($r=0.463$, $p<0.01$), and a moderately strong one with social identity ($r=0.298$, $p<0.01$).

Finally, the remaining variables had the anticipated significant positive bivariate relationships with one another with varying magnitudes (normative alignment: $r=0.352$-0.632, $p<0.01$; duty to obey: $r=0.356$-0.632, $p<0.01$; social identity: $r=0.352$-0.356, $p<0.01$).

Variable	*Treatment*	*Procedural justice*	*Normative alignment*	*Duty to obey*
Procedural justice	-0.103*			
Normative alignment	-0.114*	0.689**		
Duty to obey	-0.144**	0.463**	0.632**	
Social identity	-0.150*	0.298**	0.352**	0.356**

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

Table 2/a Correlational results

Appendix/B – Forest plots:

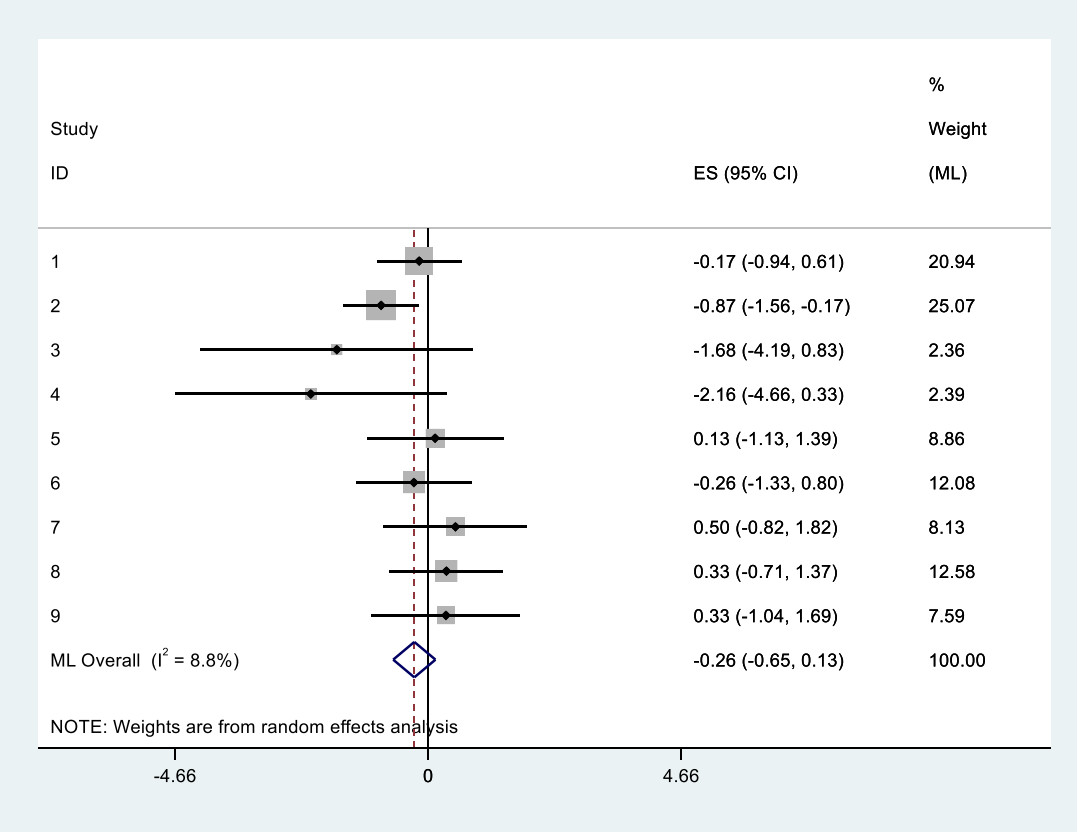


Figure 2/a Treatment effect consistency for normative alignment

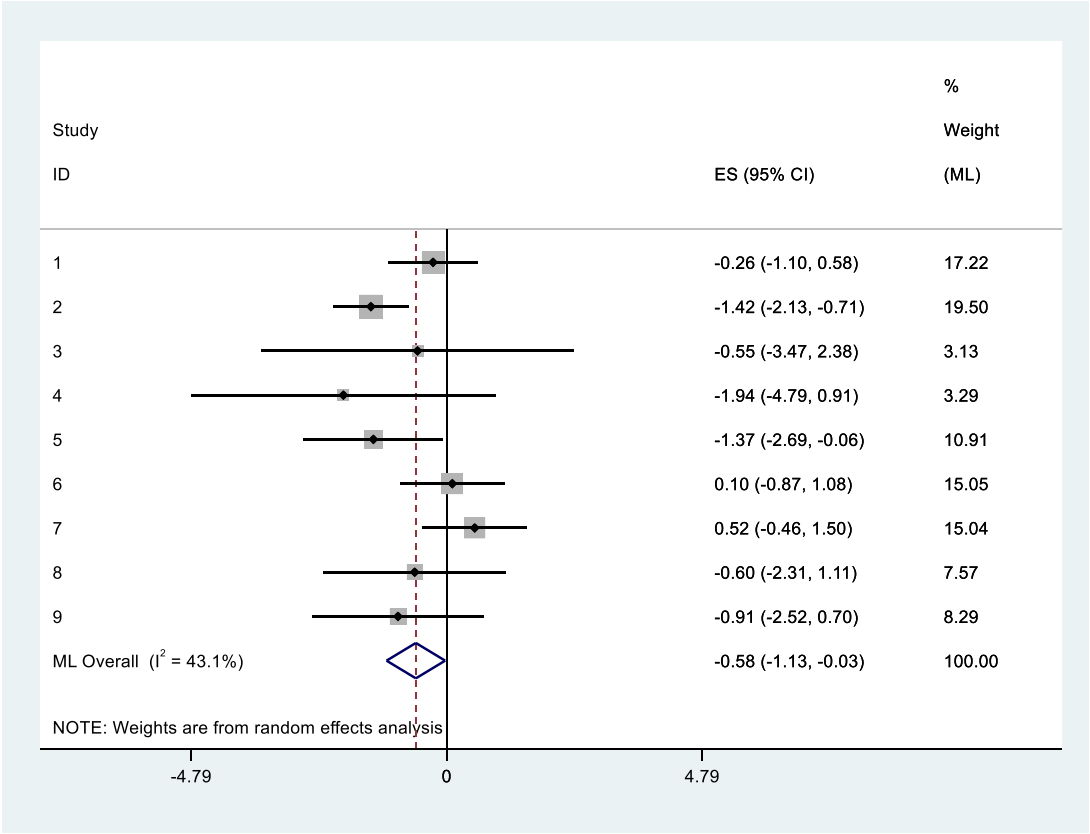


Figure 3/a Treatment effect consistency for duty to obey

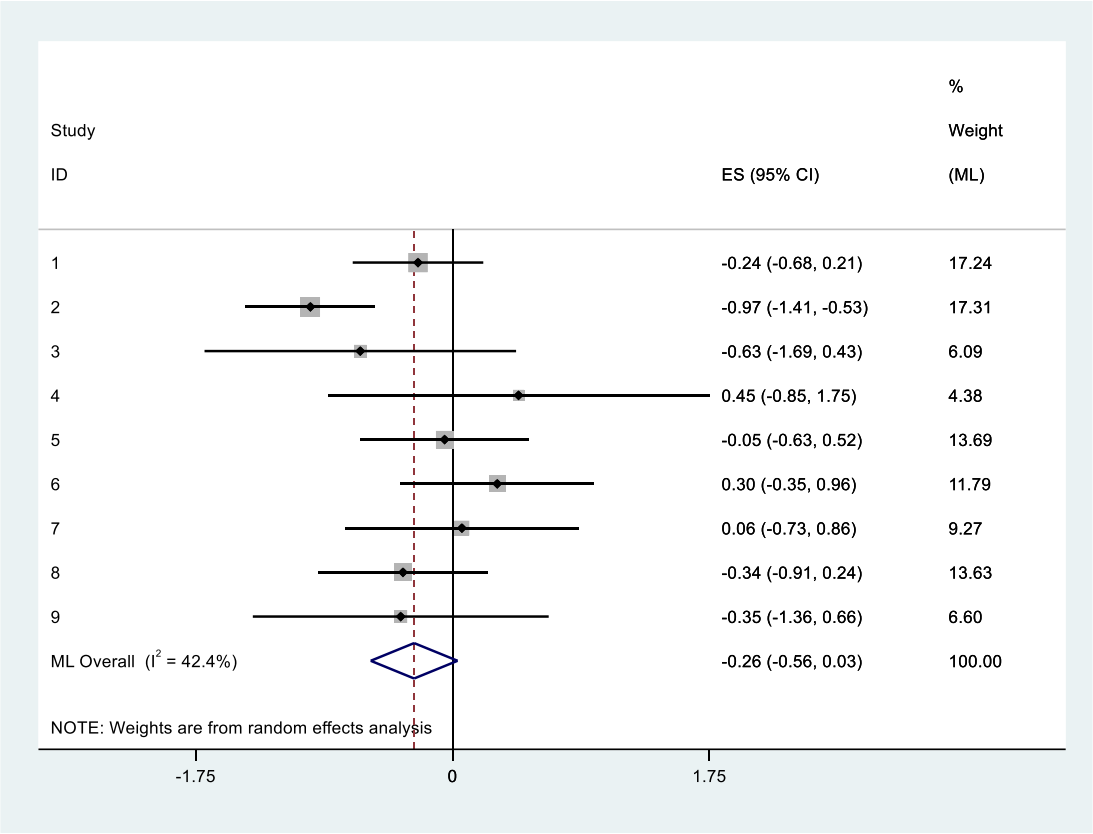


Figure 4/a Treatment effect consistency for social identity

Appendix/C – Causal mediation analysis results without covariates

Procedural justice as mediator	*Type*	*Average effect*	*Mediate %*	*Mean ρ*	*Residual R^2*	*Total R^2*
Normative alignment	NIE_{mean}	-0.247* [-0.445, -0.067]	81.1%	0.6	0.36	0.21
	NDE_{mean}	-0.047 [-0.292, 0.207]		~0.1	0.01	~0.01
Duty to obey	NIE_{mean}	-0.179* [-0.325, -0.038]	44.2%	0.5	0.25	0.19
	NDE_{mean}	-0.223 [-0.493, 0.052]		0.5	0.25	0.19
Social identity	NIE_{mean}	-0.071* [-0.133, -0.012]	24.9%	0.3	0.09	0.07
	NDE_{mean}	-0.209* [-0.384, -0.036]		0.8	0.64	0.55

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

Table 3/a Causal mediation analysis results without accounting for the pre-treatment covariates

Acknowledgments

I would like to thank Jonathan Jackson for many insightful comments and suggestions for an earlier version of this paper. I would like to also thank Sarah MacQueen and Ben Bradford for providing the dataset for the analysis.

References

Baron, Reuben M. and David A. Kenny. 1986. "Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6):173–82.

Bradford, Ben. 2014. "Policing and Social Identity: Procedural Justice, Inclusion and Cooperation between Police and Public." *Policing and Society* 24(1):22–43.

Bradford, Ben. 2017. *Stop and Search and Police Legitimacy*. Routledge.

Bradford, Ben, Aziz Huq, Jonathan Jackson, and Benjamin Roberts. 2014. "What Price Fairness When Security Is at Stake? Police Legitimacy in South Africa." *Regulation and Governance* 8(2):246–68.

Bullock, John G., Donald P. Green, and Shang E. Ha. 2010. "Yes, but What's the Mechanism? (Don't Expect an Easy Answer)." *Journal of Personality and Social Psychology* 98(4):550–58.

Coffman, D. L. and W. Zhong. 2012. "Assessing Mediation Using Marginal Structural Models in the Presence of Confounding and Moderation." *Psychological Methods* 17(4):642–64.

Cox, M. G., Y. Kisbu-Sakarya, M. Mio evi, and D. P. MacKinnon. 2013. "Sensitivity Plots for Confounder Bias in the Single Mediator Model." *Evaluation Review* 37(5):405–31.

Daniel, R. M., B. L. De Stavola, S. N. Cousens, and S. Vansteelandt. 2015. "Causal Mediation Analysis with Multiple Mediators." *Biometrics* 71(1):1–14.

Ding, Peng and Tyler J. Vanderweele. 2016. "Sharp Sensitivity Bounds for Mediation under Unmeasured Mediator-Outcome Confounding." *Biometrika* 103(2):483–90.

Fagan, Abigail A. 2017. "Illuminating the Black Box of Implementation in Crime Prevention." *Criminology & Public Policy* 16(2):451–55.

Famega, Christine, Joshua C. Hinkle, and David Weisburd. 2017. "Why Getting Inside the 'Black Box' Is Important." *Police Quarterly* 20(1):106–32.

Guyatt, Gordon H. et al. 2011. "GRADE Guidelines: 7. Rating the Quality of Evidence - Inconsistency." *Journal of Clinical Epidemiology* 64(12):1294–1302.

Haberman, Cory P. 2016. "A View inside the 'Black Box' of Hot Spots Policing from a Sample of Police Commanders." *Police Quarterly* 19(4):488–517.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American*

Statistical Association 81(396):945–60.

Hough, Mike, Jonathan Jackson, and Ben Bradford. 2013. "Legitimacy, Trust and Compliance: An Empirical Test of Procedural Justice Theory Using the European Social Survey." Pp. 326–53 in *Legitimacy and Criminal Justice - An International Exploration*, edited by J. Tankebe and A. Liebling. Oxford University Press.

Huq, A. Z. Aziz H., J. Jackson, and R. J. Trinker. 2017. "Legitimating Practices: Revisiting the Predicates of Police Legitimacy." *British Journal of Criminology* (57):1101–22.

Imai, K., D. Tingley, and T. Yamamoto. 2013. "Experimental Designs for Identifying Causal Mechanisms." *Journal of the Royal Statistical Society Series A-Statistics in Society* 176(1):5–51.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological Methods* 15(4):309–34.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105(4):765–89.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010. "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science* 25(1):51–71.

Imai, Kosuke and Marc Ratkovic. 2013. "Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation." *Annals of Applied Statistics* 7(1):443–70.

Imai, Kosuke and Teppei Yamamoto. 2013. "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments." *Political Analysis* 21(2):141–71.

Jackson, Jonathan. 2018. "Norms, Normativity, and the Legitimacy of Justice Institutions: International Perspectives." *Annual Reveiw of Law and Social Sciences* 14 In press.

Jo, Booil. 2008. "Causal Inference in Randomized Experiments With Mediational Processes." *Psychological Methods* 13(4):314–36.

Judd, Charles M. and David A. Kenny. 1981. "Process Analysis - Estimating Mediation in Treatment Evaluation." *Evaluation Reviews* 5:602–19.

Keele, Luke. 2015. "The Statistics of Causal Inference: A View from Political Methodology." *Political Analysis* 23(3):313–35.

Keele, Luke, Dustin Tingley, and Teppei Yamamoto. 2015. “Identifying Mechanisms behind Policy Interventions via Causal Mediation Analysis.” *Journal of Policy Analysis and Management* 34(4):937–63.

Kline, Rex B. 2015. “The Mediation Myth.” *Basic and Applied Social Psychology* 37(4):202–13.

Kontopantelis, Evangelos and David Reeves. 2010. “Metaan: Random-Effects Meta-Analysis.” *The Stata Journal* 10(3):395–407.

Lepage, B., D. Dedieu, N. Savy, and T. Lang. 2016. “Estimating Controlled Direct Effects in the Presence of Intermediate Confounding of the Mediator–outcome Relationship: Comparison of Five Different Methods.” *Statistical Methods in Medical Research* 25(2):553–70.

Mackinnon, David P. 2008. *Introduction to Statistical Mediation*. Erlbaum.

Mackinnon, David P., Yasemin Kisbu-sakarya, and Amanda C. Gottschall. 2013. “Developments in Mediation Analysis Oxford Handbooks Online Developments in Mediation Analysis.” Pp. 1–28 in *Oxford Handbook of Quantitative Methods*, vol. 2, edited by T. D. Little. New York: Oxford University Press.

Mackinnon, David P. and Angela G. Pirlott. 2015. “Statistical Approaches for Enhancing Causal Interpretation of the M to Y Relation in Mediation Analysis.” *Personality and Social Psychology Review* 19(1):30–43.

MacQueen, Sarah and Ben Bradford. 2015. “Enhancing Public Trust and Police Legitimacy during Road Traffic Encounters: Results from a Randomised Controlled Trial in Scotland.” *Journal of Experimental Criminology* 11(3):419–43.

MacQueen, Sarah and Ben Bradford. 2017. “Where Did It All Go Wrong? Implementation Failure—and More—in a Field Experiment of Procedural Justice Policing.” *Journal of Experimental Criminology* 13(3):321–45.

Manski, Charles F. 2007. *Identification for Prediction and Decision*. Harvard University Press.

Mazerolle, Lorraine, Emma Antrobus, Sarah Bennett, and Tom R. Tyler. 2013. “Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice.” *Criminology* 51(1):33–63.

Moerkerke, Beatrijs, Tom Loeys, and Stijn Vansteelandt. 2015. “Structural Equation Modeling versus Marginal Structural Modeling for Assessing Mediation in the Presence of Posttreatment Confounding.” *Psychological Methods* 20(2):204–20.

Murphy, Kristina and Adrian Cherney. 2012. “Understanding Cooperation with Police in a Diverse Society.” *British Journal of Criminology* 52(1):181–201.

Murphy, Kristina and Tom R. Tyler. 2017. “Experimenting with Procedural Justice Policing.” *Journal of Experimental Criminology* (August):1–6.

Muthen, B. and T. Asparouhov. 2015. “Causal Effects in Mediation Modeling: An Introduction With Applications to Latent Variables.” *Structural Equation Modeling-a Multidisciplinary Journal* 22(1):12–23.

Na, Chongmin, Thomas A. Loughran, and Raymond Paternoster. 2015. “On the Importance of Treatment Effect Heterogeneity in Experimentally-Evaluated Criminal Justice Interventions.” *Journal of Quantitative Criminology* 31(2):289–310.

Nagin, Daniel S. and Cody W. Telep. 2017. “Procedural Justice and Legal Compliance.” *Annual Review of Law and Social Science* 13(1):5–28.

Pearl, Judea. 2001. “Direct and Indirect Effects.” *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* 411–20.

Pirlott, Angela G. and David P. Mackinnon. 2016. “Design Approaches to Experimental Mediation ☆.” *Journal of Experimental Social Psychology* 66:29–38.

Rhodes, Kirsty M., Rebecca M. Turner, and Julian P. T. Higgins. 2016. “Empirical Evidence about Inconsistency among Studies in a Pair-Wise Meta-Analysis.” *Research Synthesis Methods* 7(4):346–70.

Robins, James M. and Sander Greenland. 1992. “Identifiability and Exchangeability for Direct and Indirect Effects.” *Epidemiology* 3(2):143–55.

Spencer, Steven J., Mark P. Zanna, and Geoffrey T. Fong. 2005. “Establishing a Causal Chain: Why Experiments Are Often More Effective than Mediational Analyses in Examining Psychological Processes.” *Journal of Personality and Social Psychology* 89(6):845–51.

Sunshine, Jason and Tom R. Tyler. 2003. “The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing.” *Law and Society Review* 37(3):513–48.

Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. 2014. “Mediation: R Package for Causal Mediation Analysis.” *Journal of Statistical Software* 59(5):1–38.

Trinkner, Rick and Tom R. Tyler. 2016. “Legal Socialization : Coercion versus

Consent in an Era of Mistrust." *Annual Review of Law and Social Science* 12:417–39.

Tyler, Phillip Atiba Goff, and Robert J. MacCoun. 2015. "The Impact of Psychological Science on Policing in the United States: Procedural Justice, Legitimacy, and Effective Law Enforcement." *Psychological Science in the Public Interest* 16(3):75–109.

Tyler, T., J. Fagan, and A. Geller. 2014. "Street Stops Police Legitimacy: Teachable Moments in Young Urban Men's Legal Socialization." *Journal of Empirical Legal Studies* 11(14):751–85.

VanderWeele, Tyler J. 2015. *Explanation in Causal Inference - Methods for Mediation and Interaction*. Oxford University Press.

VanderWeele, Tyler J. and Stijn Vansteelandt. 2014. "Mediation Analysis with Multiple Mediators." *Epidemiologic Methods* 2(1):95–115.

Walters, G. D. 2017. "Beyond Dustbowl Empiricism: The Need for Theory in Recidivism Prediction Research and Its Potential Realization in Causal Mediation Analysis." *Criminal Justice and Behavior* 44(1):40–58.

Walters, Glenn D. 2015. "Early Childhood Temperament , Maternal Monitoring , Reactive Criminal Thinking , and the Origin(s) of Low Self-Control." *Journal of Criminal Justice* 43(5):369–76.

Walters, Glenn D. and Jon T. Mandracchia. 2017. "Testing Criminological Theory through Causal Mediation Analysis: Current Status and Future Directions." *Journal of Criminal Justice* 49:53–64.

Interlude 1

Does procedural justice mediate the impact of previous contact on legitimacy? In short, it does appear to do so, but to a varying extent and robustness depending on which aspect of legitimacy one focusses. In the analysis of Paper 1, the effect of the treatment on normative alignment with the police was fully mediated by procedural justice, but only partially mediated on duty to obey the police. Crucially, procedural justice mediated the impact of contact on both aspects of legitimacy, regardless of where the interaction effect had been assigned. The stronger mediated effect for normative alignment implies that procedurally just cues become especially important when evaluating whether the police act in a morally appropriate manner. Procedural justice remained important when considering consent to police actions, but the partial mediation indicates that other deliberations regarding appropriate police behaviour may also play a role. Overall, these results provide causal evidence for the mediating role of procedural justice with moderately strong effect sizes and reasonable robustness. In this thesis, Paper 4 builds on these results when it examines to what degree (and which aspect of) police legitimacy mediates the impact of procedural justice on willingness to cooperate with the police.

Does procedural justice also mediate the impact of contact on psychological processes, specifically on social identification? The results of this are mixed. Although the indirect effect was significant, the effect size was very weak and sensitive compared to the other two indirect effects. Even more worryingly, the significance of the result was dependent on the assignment of the interaction effect, as the indirect effect did not differ from zero when the effect of the interaction was fully attributed to the direct effect. It is in this light that Paper 2 undertakes a more detailed examination of psychological processes (including social identification) which potentially mediate the impact of procedural justice on the legitimacy of the police and the law.

Paper 2: “It’s nice to be empowered”: An experimental assessment of psychological drivers of police legitimacy

Krisztián Pósch

Abstract

This paper considers the psychological drivers of police and legal legitimacy. Social identification, personal sense of power, the police’s grip on power, and self-control are assessed as potential mediators of the impact of procedural justice on legitimacy of the police and the law (measured as normative alignment and duty to obey). Procedurally just, procedurally unjust, and breach of boundaries conditions are compared in two randomised experiments and one experiment with parallel (encouragement) design. Statistical (Study 1 and Study 2) and design-based (Study 3) causal mediation analysis are applied to test the mediated effects in crowdsourced samples from the US and the UK (from Amazon Turk and Prolific Academic). In all three studies, only personal sense of power mediated the impact of procedural justice, and only towards normative alignment with the police and the law. Neither social identification, nor the police’s grip on power, nor self-control had a causally mediated impact. This article (1) identifies empowerment as the mediator of the impact of procedural justice on one aspect of legitimacy, (2) discusses the psychological significance of the findings, and (3) recommends and demonstrates the application of causal mediation analysis for studying indirect effects.

Keywords: causal mediation analysis, parallel design, police grip on power, police legitimacy, procedural justice, respect for boundaries, self-control, sense of power, social identity

Introduction

In the procedural justice literature, it is often posited that fair treatment by the police increases value-driven self-regulation (Hough 2012; Jackson 2018; Tyler 2009). Legitimacy strengthens self-regulation by enhancing the belief that it is the "right thing to do" to obey the law and permits a shift towards consensus-based policing tactics, which are less costly and easier to maintain than coercive strategies (Jackson and Gau 2015; Tyler, Goff, and MacCoun 2015; Tyler and Jackson 2013). People recognise that legal authorities are entitled to be obeyed when they believe that those authorities are moral, just, and appropriate. Legitimacy-driven self-regulation can inspire a sense of social responsibility, strengthen community engagement, and motivate the individual to proactively help the authorities through cooperation and reactively show deference to police orders and compliance with established rules (Jackson 2018; Tyler and Jackson 2014).

Yet, we have only a limited understanding of which psychological mechanisms are involved in this self-regulation, and more importantly, which of these mechanisms transmit the impact of procedural justice on legitimacy of the police and the law. The simplest account refers to the unmediated effect of procedural justice on legitimacy, i.e., that procedural justice is a societal expectation about how officers should wield their power and authority, and when officers are seen to act in procedurally just ways, this legitimates the institution that imbues them with that power and authority. But a key feature of procedural justice theory is that procedural justice fosters a sense of status, value, and inclusion. There is, however, no causal evidence for the existence of these effects.

This article contributes to the literature in three ways. First, it experimentally manipulates the perception of procedural justice and respect for boundaries comparing procedurally just, procedurally unjust, and breach of boundaries conditions. It thus adds to a growing (but still small) body of experimental work in this area of research, adding a focus on the relatively new concept of bounded authority. Second, this is the first paper to address the causal mechanisms through which procedural justice influences legitimacy. In particular, this paper discusses and assesses four potential candidates, social identification, personal sense of power, police grip on power, and self-control as mediators that could conceivably transmit the impact of procedural justice on the legitimacy of the police and the law. Finally, as a methodological novelty in the study of procedural justice, statistical (i.e., causal mediation analysis with post-

treatment confounding) and design-based (i.e., parallel [encouragement] design) approaches are taken, to estimate causally mediated effects and to test the robustness of these effects with sensitivity analysis techniques.

Appropriate police behaviour and legitimacy of the police and the laws

The popular conception of appropriate police behaviour is described by several expectations (e.g., police effectiveness, distributive justice), but in most Western countries procedural justice has been found to be the most influential (Tyler et al. 2015; Tyler and Jackson 2013). Procedural justice refers to individual expectations of being treated with dignity and respect, being allowed to voice one's opinions, and receiving judgments made fairly and neutrally by the police officers. A wide range of literature has demonstrated that procedural justice is a key societal norm dictating appropriate police behaviour, and as such, is a strong precondition of legitimacy (e.g., Bradford 2014; Bradford, Milani, and Jackson 2017; Hamm, Trinkner, and Carr 2017; Hough, Jackson, and Bradford 2013; Jackson et al. 2012; Mazerolle et al. 2013; Tyler and Jackson 2014).

Recent work has argued that the influence of procedurally just treatment is not without constraints, and that respect for boundaries is also a crucial element of views regarding appropriate police behaviour (Huq, Jackson, and Trinker 2017; Trinkner, Jackson, and Tyler 2017; Trinkner and Tyler 2016). Respect for boundaries represents the limits that people place on where and to what extent they accept authorities exerting their power. While procedural justice is primarily concerned with how people are treated, boundaries entail whether such treatment infringes on established rules, for instance by encroaching on parts of citizens' lives where its presence is unwanted or unwarranted. Overall, the perceived abuse and misuse of police power represent a breach of such legal boundaries. Research so far has found that procedural justice and respect for boundaries are highly correlated but separate constructs. In this paper, both procedural justice and respect for boundaries are manipulated in tandem by using three experimental scenarios of procedural justice, procedural injustice, and breach of boundaries (which assumes procedural injustice). As such, the focus is not on disentangling the potential effects of procedural justice and boundaries, but on what happens when one adds breach of boundaries to a procedurally unjust scenario.

Legitimacy is a quality which encompasses rightful power and the ensuing internalisation and willing deference to the police and the law (Jackson and Gau 2015).

This rightful power is best described as normative alignment, and willing deference as duty to obey. Normative alignment with the police represents normative justifiability of the power of the police in the eyes of the citizens; people feel that they have a shared sense of right and wrong and a common morality with the officers because, they believe, officers act in normatively appropriate ways. By contrast, free duty to obey bestows the police with the authority to command appropriate behaviour even when one disagrees with the received instructions (Hough et al. 2013; Tyler and Jackson 2014).

Compared to police legitimacy, legal legitimacy has received less attention in the literature. Legal legitimacy captures the extent to which people are ready to endorse the prevailing regulations, defer to the legal authorities, and allow laws to prescribe appropriate behaviour. Normative alignment with the law includes views regarding the normative appropriateness of the legal institutions and the laws which guide them. By contrast, duty to obey the law embodies the acceptance that the legal authorities can rightfully dictate appropriate behaviour, even if one disagrees with the substance of the law (Huq et al. 2017; Jackson and Gau 2015; Trinkner et al. 2017).

It has been found that the police and legal duty to obey are associated with reactive outcomes, such as legal compliance (Jackson 2018; Jackson et al. 2012), whilst normative alignment with the police and the law tend to be associated with proactive outcomes, especially increased community engagement and cooperation with the police (Moravcová 2016; Tyler et al. 2015; Tyler and Jackson 2014). This gives further justification to the distinction as it implies that the two aspects of legitimacy might have different motivational bases.

Psychological drivers

One of the prime candidates among the psychological drivers of police and legal legitimacy is social identification as theorised by the group-engagement model (Blader and Tyler 2009; Tyler and Blader 2003). The group-engagement model posits that judgements regarding the procedural justice of the power-holder precede and influence judgments about one's identity and that this identity mediates the effect of procedural justice towards the perception of legitimacy and behavioural outcomes. The police, as the most visible agents of the justice system, are usually considered prototypical representatives of the state and harbingers of law-abiding citizenship. In the UK context, for instance, the police are national "condensation symbols" which help

members of society articulate their collective identities (e.g., the famous "bobbies" as a symbol of Britishness). Symbols of the police and policing have become identified with the meaning of citizenship (Loader 2006).

Therefore, according to the group-engagement model, police activity emanates identity-relevant information which shapes other people's social identification to the superordinate group they represent. Specifically, when people are allowed a voice and treated with respect and dignity by the police, it shows that they are valued members of this shared superordinate group, and hence, fosters social inclusion and a sense of identification. Procedural justice cues make individuals feel that they can be proud to be a member of the group they identify with and that they have secured a certain social standing or status, which in turn engenders legitimation of and cooperation with the authorities. By contrast, unfair treatment and breach of legal boundaries signal social exclusion and diminished status which can undermine the social identification of individuals and lead to social marginalisation (Bradford 2014; Bradford et al. 2017; Meares 2017; Murphy et al. 2017; Radburn and Stott 2018).

Despite the compelling theoretical arguments, so far only a limited amount of research has addressed whether social identification mediates the impact of procedural justice on legitimacy. Results from a quasi-representative survey in England and Wales (Bradford et al. 2017), a large longitudinal survey in Australia (Bradford, Murphy, and Jackson 2014), and a randomised control trial in Scotland (Bradford et al. 2015) all found that social identification channels procedural justice's effect on police legitimacy. By contrast, Bradford's (2014) study of hard-to-reach young people in London did not find any downstream effects predicted by the theory. Unfortunately, a considerable limitation of the existing literature is that the overwhelming majority of the discussed studies are cross-sectional, and even the two which used longitudinal analysis (Bradford et al. 2014; Murphy, Bradford, and Jackson 2016) did not permit drawing causal inference.

The second psychological process that potentially mediates the impact of procedural justice on police legitimacy is sense of power, which also originates in the psychological literature: Mentovich (2012) argues that in addition to increased social identification and enhanced status, procedural justice also augments the individual's sense of control. Markedly, this is only a perception of control, which is informed by fair treatment by the power holder. When officers explain themselves, listen to concerns, and treat people with respect, those citizens feel more powerful, even if they

do not actually possess increased power over the police or have an enhanced influence concerning the outcome. To put it another way, people who are treated fairly will feel empowered compared to those who are treated unfairly. This perceived increase in mastery (i.e., a perception of one's capacity to influence others) has been referred to as control (Mentovich 2012; Ratcliff and Vescio 2017), power (Anderson, John, and Keltner 2012; Gan, Heller, and Chen 2018), or autonomy (van Prooijen 2009) in the literature, but they all refer to the same concept, which will be referred to as sense of power for the rest of the paper. Because power is inherently relational, in this paper, personal sense of power is defined as the expectant beliefs regarding one's ability to influence the police during potential future encounters (Anderson et al. 2012).

A third mediator to consider is the police's perceived grip on power. Unlike sense of power, the police's grip on power concerns the extent to which people consider the police to have power over them and in their community. As emphasised by Jackson and Gau (2015), procedural justice is not only a property of an institution, but also carries a relational quality, and describes the connection and power-differential between the authorities and citizens. It has been shown that procedural justice becomes more important when power-differences become salient, and it has only a limited or even the opposite effect in non-hierarchical settings (Mentovich 2012; van Prooijen, van den Bos, and Wilke 2002). This implies that if empowerment (i.e., increased personal sense of power) is accompanied by the experience of a reduced police grip on power, it might be detrimental to the goals of what it wants to achieve. Moreover, the perception of a reduced grip on power could also alarm the police who might get concerned with losing their clout over people (Tyler et al. 2015). Even though it appears unlikely that grip on power would mediate the impact of procedural justice on legitimacy, there are two reasons to include it in the analysis: (1) it allows us to test the impact of other mediators, conditional on police's perceived grip on power, and (2) it can inform research on whether people attribute more or less power to the police when they expect to be treated with procedural fairness.

Finally, the fourth mediator to examine is self-control. Self-control is probably best described as the cognitive capacity of an individual to control one's behavioural inhibitions (e.g., to overcome temptation) and strengthen one's determination (e.g., to reach certain goals) (Light, Rios, and DeMarree 2018; Mooijman et al. 2017). In criminology, self-control is most often connected to deviant and criminal behaviour. It has been found that people with higher levels of self-control are more likely to

comply with the law (Reisig, Tankebe, and Mesko 2014; Vazsonyi et al. 2016; Vazsonyi, Mikuška, and Kelley 2017). As with police grip on power, the expectation regarding self-control is that it is not affected by procedural justice, and it does not carry the impact of procedural justice on legitimacy of the police or the law. However, the inclusion of self-control could still be important to make the other mediators' effects conditional on the cognitive capacity to control one's behaviour.

Causal mediation analysis with multiple mediators – Study 1 and Study 2

Causal mediation analysis estimates natural direct effects (NDE) and natural indirect effects (NIE). The NDE considers the average change in the outcome when someone receives the treatment instead of the control while holding the mediator constant according to its naturally occurring value in the control group (i.e., the treatment's effect that does not involve the mediator). By contrast, the NIE holds the treatment constant as if everyone was assigned to the treatment group, and gauges the expected change in the outcome given the mediator's naturally occurring values in the control and treatment condition (i.e., the treatment's effect that goes through the mediator). This formulation of direct and indirect effects has several advantages, including non-parametric identification, the effortless incorporation of treatment-mediator interaction effects, and clearly spelt out causal identification assumptions (Imai et al. 2011; Imai, Keele, and Yamamoto 2010; Pearl 2001; VanderWeele 2016).

In the case of a single causal mediator, one of these causal identification assumptions states that, in the presence of post-treatment confounding, natural effects cannot be estimated. In other words, there cannot be other variables other than the mediator that have been affected by the treatment and would alter the mediator-outcome relationship. Thus, in essence, this assumption rules out the presence of other mediators, unless they are completely independent (i.e., orthogonal) of one another. Although it is conceivable to find such cases (e.g., Taguri, Featherstone, and Cheng 2018), for most social science examples this is very unlikely to hold true. For example, in this article, it is very implausible that views about personal sense of power and the police's grip on power would be entirely unrelated to each other.

Thus, for multiple mediators, a new set of identifying assumptions are necessary. Several solutions have been offered depending on the type of model (Steen et al. 2017; Tchetgen Tchetgen and VanderWeele 2014; VanderWeele and Vansteelandt 2014), here the one proposed by Imai and Yamamoto (2013; also see:

Keele, Tingley, and Yamamoto 2015) is discussed which suits the current linear modelling strategy the best (for an alternative with g-computation see: De Stavola et al. 2015). This method's set of (sequential ignorability) assumptions requires that controlling for a vector of pre-treatment confounders there is no unmeasured confounding for the following relationships:

1. Treatment-outcome, treatment-mediator, and treatment-post-treatment confounders (i.e., alternative mediators)
2. Mediator-outcome, also taking into account the treatment and post-treatment confounders
3. Post-treatment confounders-outcome, also taking into account the treatment

…and in addition, that:

4. There is no unmeasured post-treatment confounder that could have an impact on the mediator-outcome relationship

Crucially, from these four assumptions, only the first is satisfied by the random assignment of the treatment. Moreover, the last assumption still maintains that there cannot be any other post-treatment confounder (i.e., mediator) that has not been taken into account but would have an impact on the mediator-outcome relationship. Finally, to make the NDE and NIE estimable, a couple of parametric assumptions need to be met. First, the linearity assumption is needed for the additivity of the effects (i.e., so the NDE and NIE add up to the total effect). Second, a loosened version of effect homogeneity (Robins and Greenland 1992) has to be upheld, which asserts that on average there cannot be a treatment-mediator interaction which would affect the relationship between the mediator and the outcome.

To estimate the natural effects, a semi-parametric structural equation model is fitted which (1) allows the presence of multiple causally dependent mediators, (2) estimates heterogeneous treatment effects[5], and (3) includes an interaction between the treatment and each of the mediators (Imai and Yamamoto 2013). Figure 1 takes personal sense of power as the mediator of interest and exemplifies how the estimated pathways contribute to the NDE and NIE. At the top, the NIE incorporates the treatment's and all mediators' effects that go through the mediator of interest (here: personal sense of power) towards the outcomes. In contrast, in the middle, the NDE

[5] Unlike the traditional SEM framework where the unit homogeneity of treatment effects is assumed.

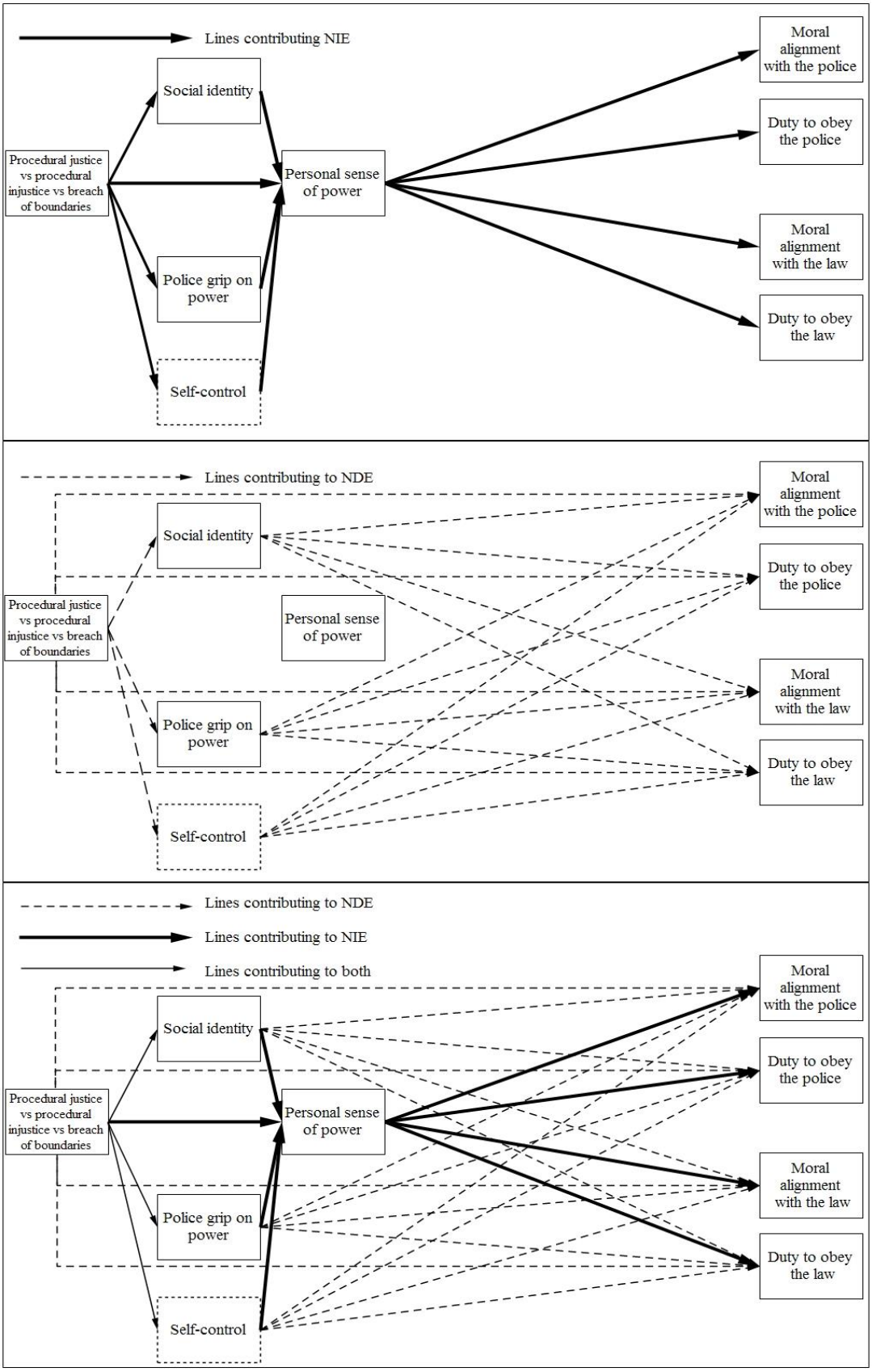


Figure 1 The estimated NDEs and NIEs for Study 1 and Study 2 with personal sense of power as an example

encompasses all pathways from the treatment and the alternative mediators that do not go through the mediator of interest. The bottom picture synthesises the two other ones by depicting all pathways. In practice, when multiple mediators are present, the mediator and outcome of interest are "rotated" and the natural effects are estimated one at a time. The mediation R package was used to estimate the NDEs and NIEs for the two randomised experiments in Study 1 and Study 2 (Tingley et al. 2014). To estimate standard errors, for each model 1000 bootstrap samples were specified.

Sensitivity analysis

The aim of sensitivity analysis is to examine and quantify the degree to which certain identifying assumptions need to be violated for the results to be dismissed as inconclusive. In this paper, two assumptions are tested with regards to the causally mediated effects with post-treatment confounding: their robustness (1) to unmeasured confounding and (2) to the presence of a treatment-mediator interaction which has an impact on the outcome. Both of these approaches are capable of detecting the sensitivity of the results, even in the presence of post-treatment confounding.

The potential sensitivity of unmeasured (pre-treatment) confounding can be gauged by taking the error terms from the model for the outcome and mediator and allowing them to be correlated. Crucially, these error terms incorporate the influence of potential omitted variables, which means that the correlation coefficient ρ (rho) captures the magnitude the correlation between the error terms needs to take to reduce the NIE to zero. Hence, a higher correlation coefficient will imply less sensitive results. A potentially easier way to interpret these ρ values is to take their squared transformation, thus creating an R^2-coefficient, which will refer to the residual variation that would need to be explained by the unmeasured confounder to nullify the results (De Stavola et al. 2015).

The second sensitivity analysis tests the potential influence of a treatment-mediator interaction that has an impact on the outcome. The presence of such an effect would be a violation of the parametric restrictions outlined earlier. Here the sensitivity parameter σ (sigma) encompasses the degree of heterogeneity in the interaction of the treatment and mediator (whilst also accounting for post-treatment confounding). In other words, σ captures the strength of the treatment-mediator interaction (i.e., the heterogeneity by the mediator) in the effect of the treatment on the outcome. There are two corresponding values of interest here: the value when the NIE becomes zero and

the value when the 95% confidence intervals of the NIE become zero due to this heterogeneity. For instance, it is conceivable that the NIE will never reach zero, however, the uncertainty caused by the heterogeneity can be so high that the confidence intervals will touch zero from the very beginning, thus indicating that the derived estimates are likely biased (Imai and Yamamoto 2013).

Notably, neither the ρ nor the σ coefficients are absolute values, they can only be interpreted with regards to other potential mediators given a set of covariates, or to a benchmark which has been established through rigorous testing. Nevertheless, these sensitivity parameters are still informative and worth estimating to quantify the potential violation of the modelling assumptions.

Designs to manipulate the mediator – Study 3

Due to the difficulties of satisfying the assumptions of causal mediation analysis discussed earlier, researchers have started to look for design-based alternatives. Early works in the design-based approach advocated separately manipulating the treatment and the mediator, and then considering the treatment's effect on the mediator, and the mediator's and treatment's effects on the outcome, thus establishing a "causal chain" (Bullock, Green, and Ha 2010; Spencer, Zanna, and Fong 2005). Unfortunately, this method only estimates the treatment's and mediator's individual causal effect(s), but these effects are not instructive as to whether or not the mediator transmits the treatment's effect towards the outcome (Keele 2015). The parallel and parallel encouragement designs overcome this limitation and are applicable in cases when the mediator can be (imperfectly) manipulated, and when such a manipulation only has an impact through the mediator (Imai, Tingley, and Yamamoto 2013; Pearl 2001).

In the parallel and parallel encouragement designs, half of the participants are randomly assigned to the control or to a treatment group. Then, as a second step, half of the participants also randomly receive the second experimental manipulation of the mediator and are assigned to a treatment or a control condition. This second manipulation removes all pre-treatment (e.g., demographics) and post-treatment (i.e., other mediators) confounders of the mediator and makes the sequential ignorability assumption unnecessary. Both designs' most important assumption is that regardless of how the values of the potential outcomes are realised, they will be identical. Thus, the potential outcomes are assumed to be contingent only on the values of the treatment and the mediator, but not on whether the participant received the second manipulation

(i.e., consistency assumption). It follows, that the second manipulation has to be subtle, to ensure that it only affects the outcome through the mediator, but does not have a direct effect. The main difference between the parallel and parallel encouragement design is that the former assumes that the second manipulation is successful ("perfect"), whilst the latter design is more circumspect, and assumes that the second manipulation is merely an encouragement which only has an effect on the people who take on the values according to the manipulation (i.e., compliers).

As with post-treatment confounding, the parallel design also requires an additional assumption of no treatment-mediator interaction that might affect the outcome (Robins and Greenland 1992), but only for the NIE to be point-identified. This no-interaction can be assumed for each unit, which is a strong, untestable, and often implausible assumption, but it allows the NIE to be non-parametrically identified. As an alternative, this no-interaction condition can be assumed on average, but then it requires the additional assumption of linearity (similar to the post-treatment confounding case). For this latter approach, the sensitivity analysis for interaction heterogeneity can be estimated. Finally, it is also possible to estimate the NIE without the no-interaction assumption, but this only allows for the estimation of sharp bounds (not the point estimate) separately for the control and treatment condition, which makes this an often unfeasible strategy.

In comparison, the parallel encouragement design is more cautious than the parallel design, as it assumes that the second manipulation was imperfect, and the participants were only encouraged to take on a certain value of the mediator. This assumption means that the NIE is only estimated among the compliers (i.e., those people who adhered to the manipulation), hence these effects are referred to as complier natural indirect effects (CNIE). For the CNIE to be estimable, assumptions akin to instrumental variables are needed: the exclusion restriction (i.e., no direct effect of the manipulation) and monotonicity (i.e., there are no defiers who would always take on the opposite value to what they were assigned to). Finally, and as a further limitation, even with all these assumptions, for the CNIE no point estimate can be derived, only sharp bounds separately for the control and treatment group.

As with the multiple mediator case, the mediation R package (Tingley et al. 2014) was used to estimate the effects with 1000 bootstrap samples for standard errors. For further details regarding how the effects are estimated please refer to Imai et al. (2013) and Imai and Yamamoto (2013).

Study 1

In Study 1, three different experimental conditions are compared to each other (listed in descending order of appropriateness of police behaviour): a procedurally just, a procedurally unjust, and a breach of boundaries condition. In addition, three causal pathways are examined, linking the experimental treatments to the legitimacy of the police and law: personal sense of power, police grip on power, and social identity. These alternative mediators are also post-treatment confounders of each other, in other words, these psychological drivers are assumed to be intertwined. Police and legal legitimacy each have two components, one which captures the normative appropriateness of the law/police (i.e., normative alignment), and a second, which considers consent to the actions of the police/letter of the law (i.e., duty to obey). All models are tested using semi-parametric structural equation models, in particular, causal mediation analysis with multiple mediators and post-treatment confounding.

Participants and procedure

Participants ("Turkers") for Study 1 were recruited on Amazon Mechanical Turk (https://www.mturk.com/). Only respondents from the United States were eligible to take part in the study, but other restrictions were not made. Turkers were randomly assigned to one of the three experimental conditions which described a case of a stop and search which claimed to be representative of the general police practice. In the text of the procedurally just condition, the police officers were polite, respectful, explained why they had stopped the protagonist, and allowed him to speak up. In the procedurally unjust condition, the officers were rude, impatient, yelling at the protagonist, and denied him the chance to voice his opinion. Finally, in the breach of boundaries condition, the officers pointed guns at the protagonist, handcuffed him, and threatened him not to report what had happened to him (the exact text of the manipulation is available in Appendix/A). To bring the text closer to the participant, thus potentially augmenting the effect of the manipulation (Maglio, Trope, and Liberman 2013), the story took place in the state they were from, and in particular the second largest city in that state.

To screen for satisficers, two attention checks (Hauser and Schwarz 2016) were included which asked where the protagonist in the vignette had been spending his time before his encounter with the police, and what the police officers were looking for. In addition, an instructional manipulation check (Anduiza and Galais 2016;

Oppenheimer, Meyvis, and Davidenko 2009) was also included, which requested the participants not to select a response to a particular item. Failure of either of these checks meant the end of the respondent's participation. IP-protection was used to prevent the same participant entering the study multiple times. The questions were separated into blocks based on the construct of interest, and in each block, the item order was randomised to mitigate ordering and placement effects (Malhotra 2008; Tourangeau, Couper, and Conrad 2013). The participants were forced to answer all questions which effectively resulted in no missing data. At the end of the survey the participants were debriefed, informed that the article was made up, told about the goal of the study, provided with a link to FBI statistics regarding crime rates across the US, and were allowed to withdraw their participation without risking the loss of monetary compensation ($0.5). None of them decided to do so.

Altogether 403 people finished the survey with almost the same number of participants in each experimental condition (procedurally just: n=134, procedurally unjust: n=135, breach of boundaries: n=134). In addition to state of residence, gender, age, education, ethnic minority background, political orientation, and police- and citizen-initiated contact were measured as pre-treatment covariates. Because of the small number or absence of people from certain minority groups, this categorical variable was recoded to a binary one (white vs ethnic minority background) and entered as such into future analysis. Overall, 44% of the participants were women (n=179), with the average age of 35.9, and almost half of the participants had at least a college degree (n=200). Approximately 78% of the participants were White (n=315) and on average they were slightly left-leaning in their political views (M=3.4 on a 1-7 left-right scale). 23% (n=92) of them reported that they had initiated contact with the police and 31% (n=124) that the police had contacted them in the last two years. On average, the experimental conditions were balanced across all covariates. For details regarding the distribution of the pre-treatment covariates and manipulation checks please refer to Appendix/A which contains the balance table and manipulation checks for Study 1.

Measurements

The binary treatment variable was defined by the experimental conditions being compared. Therefore, three treatment variables were derived: one for the procedurally just-procedurally unjust (Pj vs unpj), one for the procedurally just-breach of

boundaries (Pj vs bob), and one for the procedurally unjust-breach of boundaries comparison (Unpj vs bob). Accordingly, for each outcome variable, three models were fitted for each mediator depending on the conditions being juxtaposed. For all binary treatments, 1 always stood for more positive views about the police and 0 for the less favourable one in the pair.

The measures of procedural justice, respect for boundaries, police and legal legitimacy were taken from Huq, Jackson, and Trinker (2017). Personal sense of power was an adapted version of the questionnaire put forward by Anderson et al. (2012) while social identification was captured akin to the ones used by Bradford, Murphy, and Jackson (2014). Finally, questions of police grip on power were adapted based on MacQueen and Bradford (2015).

All negative items were reversed so higher numbers would indicate stronger agreement with the construct. Next, variables were entered in a confirmatory factor analysis (CFA) with each of them being determined by the latent variable of the construct they were intended to measure. The model fit indices for the CFA implied appropriate fit. The factor loadings, Cronbach's Alphas, and average inter-item correlations all substantiated the internal consistency of the various latent variables. For each scale, the confirmatory factor scores were derived and used in every subsequent analysis. For further details regarding the question and item wording, the CFA results, and other measures of internal consistency please refer to Appendix/A.

Results

In all models gender, age, level of education, ethnicity, political orientation, and previous police- and citizen-initiated contact were entered as a vector of pre-treatment covariates. For the sake of brevity, each of the three comparisons of the experimental conditions is denoted by a subscript (pjunpj: procedurally just-procedurally unjust, pjbob: procedurally just-breach of boundaries, unpjbob: procedurally unjust-breach of boundaries). As a first step, estimates of the average treatment effects (ATE; denoted by βs below) for the mediators were derived using linear regression analysis with 1000 bootstrap samples. Assessing the ATEs showed that sense of power had the strongest significant increase under all comparisons (β_{pjunpj}=0.336, $CI_{95\%}$=[0.156, 0.517], $p<0.01$; β_{pjbob}=0.744, $CI_{95\%}$=[0.574, 0.914], $p<0.01$; $\beta_{unpjbob}$=0.418, $CI_{95\%}$=[0.229, 0.606], $p<0.01$). Grip on power was also significantly boosted by the treatment by all but one comparison (β_{pjunpj}=0.114, $CI_{95\%}$=[0.019, 0.208], $p<0.05$; β_{pjbob}=0.186,

$CI_{95\%}$=[0.088, 0.283], p<0.01; $\beta_{unpjbob}$=0.079, $CI_{95\%}$=[-0.010, 0.168], p>0.05). By contrast, the ATE for social identity was only significant under one comparison (β_{pjunpj}=0.090, $CI_{95\%}$=[-0.073, 0.254], p>0.05; β_{pjbob}=0.220, $CI_{95\%}$=[0.060, 0.381], p<0.01; $\beta_{unpjbob}$=0.120, $CI_{95\%}$=[-0.058, 0.297], p>0.05).

Table 1 contains the causal mediation analysis results for the three mediators and four outcomes. For each dependent variable, there are three columns, each referring to a comparison between two experimental conditions. To exemplify how to interpret the results, take the first column in Table 1, where the treatment is the comparison between the procedurally just and procedurally unjust conditions and the outcome is normative alignment with the police. After taking pre- and post-treatment confounding into account, sense of power has a moderately strong indirect effect on normative alignment (NIE_{pjunpj}=0.315, p<0.01) with a still significant direct effect of the treatment (NDE_{pjunpj}=0.254, p<0.01). The mediated effect of sense of power is robust to interaction heterogeneity, as neither the average NIE nor its confidence intervals ever reach zero ($NIE_{pjunpj}(\sigma)$ =NA, $NIE_{pjunpj}(\sigma_{95\%})$=NA). Sense of power's NIE is also robust to unmeasured confounding: on average the error terms would need to have 0.8 correlation to nullify the effect, in other words, a hypothetical pre-treatment confounder would need to explain 64% of the residual variation to make the indirect effect zero. By contrast, neither grip on power (NIE_{pjunpj}=-0.027, p>0.05) nor social identity (NIE_{pjunpj}=-0.046, p>0.05) mediated the treatment's effect on the outcome, but the treatment still had a profound significant direct effect on normative alignment with the police (grip on power: NDE_{pjunpj}=0.610, p<0.01, social identity: NDE_{pjunpj}=0.632, p<0.01). Both grip on power ($NIE_{pjunpj}(\sigma)$=0.264, $NIE_{pjunpj}(\sigma_{95\%})$=0.000) and social identity ($NIE_{pjunpj}(\sigma)$=0.239, $NIE_{pjunpj}(\sigma_{95\%})$=0.000) appeared to be sensitive to interaction heterogeneity, with the confidence intervals of the indirect effect reaching zero from the very beginning. Likewise, both mediators were more sensitive to unmeasured confounding, than sense of power, with much smaller correlation coefficients and R^2s (grip on power: ρ_{pjunpj}=0.1, $R^2_{pjunpj\_residual}$=1%, social identity: ρ_{pjunpj}=0.3, $R^2_{pjunpj\_residual}$=9%).

The overview of the total effects implies that the comparison between the procedurally just and breach of boundaries conditions had the biggest effect, followed by the procedurally just-procedurally unjust, and the procedurally unjust-breach of boundaries comparisons. Overall, from the three mediators, only sense of power emerged with robust significant indirect effects, and only for normative alignment with

Causal mediation analysis with multiple mediators		*Normative alignment with the police*			*Obligation to obey the police*			*Normative alignment with the law*			*Obligation to obey the law*		
		Pj vs unpj	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*
Sense of power	*NIE*	0.315** [0.109, 0.521]	0.558** [0.167, 0.949]	0.358** [0.167, 0.549]	0.144 [-0.232, 0.520]	0.674 [-0.023, 1.371]	0.459** [0.123, 0.805]	0.310** [0.115, 0.505]	0.493** [0.135, 0.851]	0.208** [0.013, 0.403]	0.183 [-0.119, 0.485]	0.410 [-0.148, 0.968]	0.211 [-0.068, 0.471]
	NDE	0.254** [0.114, 0.394]	0.331** [0.162, 0.500]	0.059 [-0.058, 0.211]	0.356** [0.106, 0.606]	0.378* [0.063, 0.693]	-0.028 [-0.288, 0.232]	0.165* [0.011, 0.319]	0.331** [0.126, 0.536]	0.161 [-0.016, 0.338]	0.226* [0.002, 0.450]	0.350* [0.079, 0.621]	0.021 [-0.218, 0.261]
	NIE(σ)	*NA [NA]*	*NA [NA]*	*NA [NA]*	*0.730 [0.000]*	*NA [0.000]*	*NA [0.065]*	*NA [NA]*	*NA [NA]*	*NA [0.828]*	*NA [0.000]*	*NA [0.000]*	*NA [0.000]*
	ρ	*0.8 [64%]*	*0.8 [64%]*	*0.8 [64%]*	*0.3 [9%]*	*0.3 [9%]*	*0.3 [9%]*	*0.6 [36%]*	*0.7 [49%]*	*0.6 [36%]*	*0.3 [9%]*	*0.3 [9%]*	*0.3 [9%]*
Grip on power	*NIE*	-0.027 [-0.144, 0.090]	-0.082 [-0.257, 0.093]	0.035 [-0.087, 0.157]	-0.058 [-0.367, 0.251]	0.412 [-0.196, 1.020]	0.287* [0.004, 0.570]	-0.023 [-0.128, 0.082]	-0.123 [-0.301, 0.055]	-0.008 [-0.127, 0.111]	~ -0.007 [-0.13, 0.13]	-0.107 [-0.313, 0.099]	-0.025 [-0.146, 0.104]
	NDE	0.610** [0.400, 0.820]	1.034** [0.811, 1.257]	0.397** [0.189, 0.605]	0.687** [0.281, 1.093	0.378 [-0.308, 1.064]	-0.160 [-0.511, 0.191]	0.445** [0.251, 0.639]	0.891** [0.672, 1.110]	0.407** [0.195, 0.619	0.414** [0.172, 0.656]	0.650 [0.397, 0.903]	0.146 [-0.073, 0.368]
	NIE(σ)	*0.264 [0.000]*	*0.782 [0.000]*	*0.347 [0.000]*	*0.483 [0.000]*	*NA [0.000]*	*NA [0.371]*	*0.256 [0.000]*	*1.164 [0.000]*	*0.169 [0.000]*	*0.151 [0.000]*	*1.012 [0.000]*	*0.245 [0.000]*
	ρ	*0.1 [1%]*	*0.1 [1%]*	*0.05 [~1%]*	*0.2 [4%]*	*0.3 [9%]*	*0.07 [~1%]*	*0.03 [~1%]*	*0.1 [1%]*	*0.2 [4%]*	*0.01 [~1%]*	*0.1 [1%]*	*0.1 [1%]*

Social identity	*NIE*	-0.046 [-0.167, 0.075]	-0.133 [-0.346, 0.080]	0.027 [-0.136, 0.190]	-0.070 [-0.401, 0.261]	0.276 [-0.373, 0.925]	0.223 [-0.041, 0.487]	-0.011 [-0.110, 0.088]	-0.101 [-0.290, 0.088]	0.016 [-0.132, 0.164]	0.022 [-0.102, 0.146]	-0.074 [-0.274, 0.126]	0.008 [-0.145, 0.162]
	NDE	0.632** [0.408, 0.856]	1.215** [0.903, 1.527]	0.494** [0.236, 0.752	0.672** [0.266, 1.078]	0.437 [-0.245, 1.119]	-0.093 [-0.434, 0.248]	0.409** [0.210, 0.608]	0.926** [0.638, 1.214]	0.452** [0.209, 0.694]	0.322** [0.084, 0.560]	0.572** [0.303, 0.841]	0.175 [-0.078, 0.426]
	NIE(σ)	*0.239 [0.000]*	*0.696 [0.000]*	*0.153 [0.000]*	*0.355 [0.000]*	*NA [0.000]*	*NA [0.000]*	*0.077 [0.000]*	*0.528 [0.000]*	*0.086 [0.000]*	*0.129 [0.000]*	*0.392 [0.000]*	*0.088 [0.000]*
	ρ	*0.3 [9%]*	*0.3 [9%]*	*0.4 [16%]*	*0.3 [9%]*	*0.3 [9%]*	*0.3 [9%]*	*0.3 [9%]*	*0.3 [9%]*	*0.4 [16%]*	*0.3 [9%]*	*0.3 [9%]*	*0.3 [9%]*

**p<0.05, **<0.01, pj = procedurally just conditions, unpj = procedurally unjust condition, bob = breach of boundaries condition*

Table 1 Causal mediation analysis with multiple mediators – Study 1

the police and the law. These indirect effects (NIE=0.310-0.558, $p<0.01$) were always larger than the direct effects, sometimes fully mediating the treatment's effect (i.e., in the case of the procedurally unjust-breach of boundaries comparison). Sense of power's NIEs were also robust to interaction heterogeneity, as their average NIEs never reached zero (NIE(σ)=NA) and their confidence intervals only well beyond the 95% range, or not at all (NIE($\sigma_{95\%}$)=0.828-NA). Sense of power's NIEs for normative alignment with the police and law were also relatively robust to unmeasured confounding (ρ=0.6-0.8, $R^2_{residual}$=36-64%). All the other significant effects (i.e., for obligation to obey the police and under the procedurally unjust-breach of boundaries comparison) were either sensitive to interaction heterogeneity (sense of power: $NIE_{unpjbob}(\sigma)$=NA, $NIE_{unpjbob}(\sigma_{95\%})$=0.065) or to unmeasured confounding (grip on power: $\rho_{unpjbob}$=0.07, $R^2_{unpjbob\_residual}$=~1%). These findings show the utility of the sensitivity analyses, which can identify significant, but spurious effects.

The lack of robust significant indirect effects of either of the obligation to obey outcomes is startling, especially because the effect sizes for sense of power as mediator are relatively large. Nevertheless, the uncertainty described by the confidence intervals shows that these are all noisy estimates.

The NDEs for the procedurally just-procedurally unjust and procedurally unjust-breach of boundaries comparisons were significant almost without exception for all outcome variables, while for the procedurally unjust-breach of boundaries comparison they were significant for the two normative alignment outcomes, with grip on power and social identity as the mediators.

Discussion

Based on the results, neither the police's perceived grip on power nor social identification appear to transmit the impact of procedural justice towards legitimacy of the police or the law, after considering pre- and post-treatment confounding. By contrast, personal sense of power seems to mediate the impact of the treatment, with relatively high robustness to unmeasured confounding and interaction heterogeneity, but only towards normative alignment with the police and the law. This implies, that higher levels of procedural justice empower the individuals, who in turn, perceive police conduct and the letter of the law as more appropriate and in line with the shared values of the community.

The very sizable confidence intervals and interaction heterogeneity for the obligation to obey outcomes indicate that the mediating effects are highly dependent on the experimental condition. This emergence of uncertainty and dependency might be closely related to how these constructs are measured. The participants' willingness to obey the police or the laws despite disagreeing with them, might have different meanings when the police appear to be unjust/unlawful. Hostile environments could trigger two opposite behaviours, making certain people more likely to reject the authority of the police/law (i.e., decreased obligation to obey), whilst some others might still be willing to obey them but out of fear (i.e., increased obligation to obey). It is clear that these contradictory processes can easily result in noisy estimates, which makes the measurement of free duty to obey more difficult.

The prevalence of the direct effects shows that changing the treatment effect by comparing different levels of appropriate police behaviour is meaningful, and also that – in most cases – personal sense of power on its own cannot transmit the various treatments' entire influence. This implies that there are other psychological processes that might be in play which have not been included in Study 1. Furthermore, higher levels of appropriate police behaviour are predictive of all the legitimacy outcomes (unless their impact is fully mediated by sense of power), which supports the initial hypothesis that the expectation of procedurally just police practices increases the legitimacy of the police and the law.

Study 2

Study 2 was built on very similar premises to Study 1, with two important changes. First, instead of being crowdsourced in the US, this study was fielded in the UK. Because this paper focusses on psychological motivators of police legitimacy, these psychological processes should be very similar, at least in countries with similar cultural traditions and similar levels of economic development. Moreover, earlier observational studies showed that findings in the UK are often similar to ones found in the US (e.g., Huq et al. 2017; Jackson et al. 2012; Jackson and Sunshine 2007).

As a second change, self-control was added as a post-treatment confounder and potential fourth causal pathway. One of the core tenets of procedural justice theory is that fair and respectful treatment and impartial decision making by the police activates value-driven self-regulation in individuals, and that this increased self-regulation will be more likely to result in compliance with the law. The question still remains, whether

the increased experience of mastery over the situation during police encounters (i.e., sense of power) is actually accompanied by higher reported self-control. Because in Study 1 only sense of power emerged as a significant predictor, it is worth adding self-control to help distinguish between the two constructs, both implying a certain level of control over their life/potential police encounters. With these changes in mind, Study 2 can be considered as a partial replication and extension of Study 1.

Participants and procedure

The respondents for Study 2 were recruited on Prolific Academic, the only restriction made was that they needed to be UK residents. The same procedure was followed as described for Study 1, although there were a few slight modifications made to adapt to the change of country. First, linguistic changes were made from US to British English (e.g., rumor to rumour, apologize to apologise, and so on). Second, instead of state of origin, here region of origin and the second largest city in that region was used in the article. Third, in the text of the breach of boundaries condition, police officers had tasers instead of guns, in accordance with UK practice that the police do not carry guns. Fourth, in the debriefing, the participants were pointed to the Independent Office for Police Conduct's website instead of the FBI's. Finally, the monetary compensation was increased to £1.25, to follow Prolific's regulations.

Altogether, 323 participants took part in the experiment with approximately the same number of people in each condition (procedurally just condition: n=107, procedurally unjust condition: n=107, breach of boundaries condition: n=109). Nobody decided to withdraw his/her participation in the study. The same pre-treatment covariates were measured as in Study 1, with one addition: the respondents were asked whether they thought with hindsight that Brexit was a good decision. As with Study 1, too few respondents belonged to the various ethnic minority groups, hence ethnicity was coded as a binary variable. To capture political orientation in the UK, a component score was derived using the left-right scale and views on Brexit ($r=0.471$, $p<0.001$), and this new variable was used in all subsequent analysis. In Study 2 68% of participants were women (n=220), the average age was 35.5, with 58% of the participants having at least a college degree (n=187). The vast majority of the respondents were White (87%, n=280), on average left-of-centre (M=3.1 on a 1-7 left-right scale) and slightly apprehensive about Brexit (M=3.5 on a 1-7 very wrong-very right decision scale). Almost 30% of them were contacted by the police (n=96) and

little more than 31% had contacted the police (n=102) in the last two years. Again, the balance tests indicated that the three experimental conditions were approximately the same with regards to the pre-treatment covariates. The detailed frequencies, means, and standard deviation of the variables across the different experimental conditions can be found in Appendix/B, including the manipulation checks for Study 2.

Measurements

For the majority of the constructs, the same measurements were used as in Study 1 with slight modifications to accommodate the change in location (again, switch to British English and to the UK from US). However, there were some alterations and additions made regarding the measurement of some of the mediators. Based on the pilot study for this experiment, police grip on power received two new items, with item-specific response alternatives. Also based on the pilot, two items were removed and two new items were introduced to measure social identity, adopting measures used in Jackson et al. (2014). Finally, a new scale was included to measure self-control. All items were selected from the scale used by Reisig et al. (2014). For details regarding the changes of measurement and the question wording of the new items and construct please refer to Appendix/B. The same methods and procedures were used as in Study 1 to transform the variables, and to analyse the model fit and internal consistency of the measures, suggesting good fit and strong reliability (for details please refer to Appendix/B).

Results

As in Study 1, all pre-treatment covariates were accounted for in all models, and the same analytic strategy was pursued. The ATEs remained strong and significant for sense of power for two comparisons out of three (β_{pjunpj}=0.795p, $CI_{95\%}$=[0.586, 1.004], p<0.01; β_{pjbob}=0.959, $CI_{95\%}$=[0.769, 1.150], p<0.01; $\beta_{unpjbob}$=0.171, $CI_{95\%}$=[-0.008, 0.350], p>0.05) and for one comparison in the case of police grip on power (β=0.096$_{pjunpj}$, $CI_{95\%}$=[-0.099, 0.290], p>0.05; β_{pjbob}=0.232, $CI_{95\%}$=[0.036, 0.428], p<0.05; $\beta_{unpjbob}$=0.171, $CI_{95\%}$=[-0.009, 0.351], p>0.05). The ATEs for social identification (β=-0.036$_{pjunpj}$, $CI_{95\%}$=[-0.167, 0.096], p>0.05; β_{pjbob}=0.006, $CI_{95\%}$=[-0.130, 0.142], p>0.05; $\beta_{unpjbob}$=0.019, $CI_{95\%}$=[-0.111, 0.148], p>0.05) and self-control (β=-0.191$_{pjunpj}$, $CI_{95\%}$=[-0.435, 0.053], p>0.05; β_{pjbob}=-0.079, $CI_{95\%}$=[-0.315, 0.156], p>0.05; $\beta_{unpjbob}$=0.095, $CI_{95\%}$=[-0.116, 0.307], p>0.05) were not significant.

Causal mediation analysis with multiple mediators		*Normative alignment with the police*			*Obligation to obey the police*			*Normative alignment with the law*			*Obligation to obey the law*		
		Pj vs unpj	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*
Sense of power	*NIE*	0.593** [0.398, 0.788]	0.902** [0.670, 1.134]	0.181 [-0.009, 0.374]	0.207* [0.090, 0.324]	0.398** [0.231, 0.565]	0.079 [-0.010, 0.168]	0.432** [0.260, 0.604]	0.667** [0.427, 0.917]	0.132 [-0.014, 0.279]	0.078 [-0.036, 0.192]	-0.046 [-0.238, 0.149]	0.060 [-0.016, 0.144]
	NDE	0.283* [0.067, 0.499]	0.369** [0.135, 0.604]	0.201 [-0.003, 0.403]	0.203 [-0.009, 0.412]	0.244* [0.019, 0.470]	0.116 [-0.086, 0.319]	0.254* [0.028, 0.480]	0.423** [0.165, 0.681]	0.268* [0.019, 0.518]	0.057 [-0.208, 0.326]	0.092 [-0.196, 0.380]	0.086 [-0.164, 0.338]
	NIE(σ)	*0.776 [0.560]*	*NA [NA]*	*NA [0.282]*	*0.273 [0.150]*	*0.550 [0.357]*	*0.131 [0.000]*	*0.564 [0.377]*	*0.926 [0.651]*	*0.232 [0.000]*	*0.133 [0.000]*	*0.327 [0.103]*	*0.147 [0.000]*
	ρ	*0.7 [49%]*	*0.7 [49%]*	*0.7 [49%]*	*0.3 [9%]*	*0.4 [16%]*	*0.3 [9%]*	*0.5 [25%]*	*0.5 [25%]*	*0.4 [16%]*	*0.2 [4%]*	*0.2 [4%]*	*0.2 [4%]*
Grip on power	*NIE*	-0.012 [-0.069, 0.045]	-0.034 [-0.091, 0.024]	-0.076 [-0.160, 0.008]	-0.001 [-0.023, 0.022]	0.015 [-0.021, 0.053]	-0.003 [-0.033, 0.030]	-0.012 [-0.076, 0.056]	-0.040 [-0.105, 0.024]	-0.079 [-0.173, 0.014]	-0.024 [-0.085, 0.038]	-0.044 [-0.108, 0.020]	-0.051 [-0.113, 0.017]
	NDE	0.887** [0.661, 1.113]	1.298** [1.050, 1.548]	0.458** [0.209, 0.708]	0.410** [0.220, 0.600]	0.623** [0.422, 0.824]	0.198 [-0.016, 0.412]	0.698** [0.458, 0.938]	1.125** [0.877, 1.368]	0.479** [0.207, 0.751]	0.159 [-0.047, 0.360]	0.356** [0.122, 0.589]	0.198 [-0.049, 0.444]
	NIE(σ)	*0.100 [0.000]*	*0.100 [0.000]*	*0.142 [0.000]*	*0.110 [0.000]*	*0.114 [0.000]*	*0.147 [0.000]*	*0.117 [0.000]*	*0.132 [0.000]*	*0.184 [0.000]*	*0.121 [0.000]*	*0.132 [0.000]*	*0.164 [0.000]*
	ρ	*0.3 [9%]*	*0.2 [4%]*	*0.4 [16%]*	*0.001 [~1%]*	*0.1 [1%]*	*0.001 [~1%]*	*0.2 [4%]*	*0.2 [4%]*	*0.3 [9%]*	*0.3 [9%]*	*0.2 [4%]*	*0.2 [4%]*

Social identity	*NIE*	0.002 [-0.053, 0.057]	0.012 [-0.037, 0.061]	0.004 [-0.041, 0.045]	-0.020 [-0.111, 0.071]	0.009 [-0.082, 0.100]	0.010 [-0.072, 0.092]	-0.001 [-0.006, 0.006]	0.010 [-0.050, 0.070]	0.007 [-0.057, 0.071]	-0.021 [-0.127, 0.105]	0.012 [-0.094, 0.116]	0.012 [-0.089, 0.113]
	NDE	0.873** [0.657, 1.089]	1.254** [1.018, 1.490]	0.378* [0.095, 0.661]	0.430** [0.265, 0.595]	0.629** [0.439, 0.820]	0.185 [-0.008, 0.378]	0.685** [0.469, 0.900]	1.075** [0.827, 1.323]	0.392** [0.115, 0.671]	0.156 [-0.037, 0.349]	0.300* [0.077, 0.523]	0.134 [-0.091, 0.359]
	NIE(σ)	*0.157 [0.000]*	*0.157 [0.000]*	*0.195 [0.000]*	*0.173 [0.000]*	*0.179 [0.000]*	*0.203 [0.000]*	*0.183 [0.000]*	*0.206 [0.000]*	*0.254 [0.000]*	*0.189 [0.000]*	*0.207 [0.000]*	*0.227 [0.000]*
	ρ	*0.3 [9%]*	*0.2 [4%]*	*0.2 [4%]*	*0.4 [16%]*	*0.4 [16%]*	*0.4 [16%]*	*0.3 [9%]*	*0.3 [9%]*	*0.2 [4%]*	*0.5 [25%]*	*0.5 [25%]*	*0.4 [16%]*
Self-control	*NIE*	0.017 [-0.017, 0.051]	-0.001 [-0.043, 0.043]	0.002 [-0.023, 0.027]	-0.010 [-0.043, 0.023]	-0.007 [-0.050, 0.036]	-0.007 [-0.022, 0.008]	0.021 [-0.016, 0.058]	-0.002 [-0.042, 0.038]	-0.001 [-0.032, 0.031]	-0.025 [-0.074, 0.024]	-0.006 [-0.052, 0.041]	0.012 [-0.029, 0.053]
	NDE	0.859** [0.624, 1.094]	1.265** [1.021, 1.509]	0.380** [0.102, 0.658]	0.420 [0.229, 0.611]	0.644** [0.440, 0.848]	0.188 [-0.036, 0.412]	0.664** [0.446, 0.882]	1.087** [0.841, 1.333]	0.400** [0.104, 696]	0.160 [-0.056, 0.377]	0.318* [0.082, 0.554]	0.134 [-0.109, 0.377]
	NIE(σ)	*0.075 [0.000]*	*0.074 [0.000]*	*0.119 [0.000]*	*0.082 [0.000]*	*0.085 [0.000]*	*0.124 [0.000]*	*0.087 [0.000]*	*0.098 [0.000]*	*0.155 [0.000]*	*0.089 [0.000]*	*0.098 [0.000]*	*0.138 [0.000]*
	ρ	*0.001 [~1%]*	*0.2 [4%]*	*0.1 [1%]*	*0.1 [1%]*	*0.2 [4%]*	*0.1 [1%]*	*0.001 [~1%]*	*0.2 [4%]*	*0.1 [1%]*	*0.2 [4%]*	*0.2 [4%]*	*0.2 [4%]*

**p<0.05, **<0.01, pj = procedurally just conditions, unpj = procedurally unjust condition, bob = breach of boundaries condition*

Table 2 Causal mediation analysis with multiple mediators – Study 2

As shown in Table 2, from the four mediators only sense of power transmitted significantly the effect of the treatment towards normative alignment with the police and the law and obligation to obey the police, and only when juxtaposing the procedurally just-procedurally unjust and procedurally just-breach of boundaries conditions. These NIEs either partially or fully mediated the treatment's effect, but they were always larger than the NDEs in the corresponding model (NIE=0.207-0.902, $p<0.05$-0.01; NDE=0.203-0.423, $p>0.05$-$p<0.01$). The NIEs for the two normative alignment outcomes were less sensitive to interaction heterogeneity and unmeasured confounding (ρ=0.5-0.7, $R^2_{residual}$=25-49%, NIE(σ)=0.564-NA, NIE($\sigma_{95\%}$)=0.377-NA) than the ones for duty to obey (ρ=0.3-0.4, $R^2_{residual}$=9-16%, NIE(σ)=0.273-0.550, NIE($\sigma_{95\%}$)=0.150-0.273). Notably, the sensitivity parameters of the coefficients for the two normative alignment outcomes were slightly less robust compared to the ones in Study 1.

The total effects were the weakest when the procedurally unjust and breach of boundaries conditions were juxtaposed, which resulted in the NDEs not being significant for some of the models (especially for the two obligation to obey outcomes). The procedurally just-breach of boundaries comparison produced the strongest treatment effect, followed by the procedurally just-procedurally unjust comparison, and for these two – similar to Study 1 – the NDEs were almost always significant.

Discussion

An important difference between Study 1 and Study 2 is the diminished average treatment effect for the comparison between the procedurally unjust and breach of boundaries conditions (despite the successful manipulation, as described in Appendix/B). The lower effect might be partially due to the slight change in the text of the breach of boundaries manipulation, where the police officers pointed tasers at the protagonist instead of guns, which would be more typical in the UK context. This, however, raises the question of whether the stronger effect – and lower reported procedural justice and respect for boundaries – in Study 1 was due to the police being perceived as threatening. Otherwise, the total effects of Study 1 and Study 2 were fairly similar, with the procedurally just-breach of boundaries comparison providing the strongest, and the procedurally just-procedurally unjust comparison the second strongest effect.

Despite adding one more post-treatment confounder, sense of power still emerged as the only mediator of the impact of procedural justice on normative alignment with the police and the law, and now also obligation to obey the police. From these three outcomes, the indirect effects of the two normative alignment outcomes were much stronger and less sensitive than the ones for duty to obey the police. Notably, however, even the results for the two normative alignment outcomes were more sensitive than the ones in Study 1. This was especially true for normative alignment with the law where the interaction heterogeneity's impact nullified the confidence intervals before those took on the value of the average NIE, and an unmeasured confounder would have needed to explain only a quarter of the residual variance to make the NIE zero.

Notably, and unlike Study 1, the standard errors of the NIE estimates for the two obligation to obey outcomes were much lower, with on average diminished (but sometimes significant) effect sizes for sense of power's NIE. This means that the lack of indirect effects found in Study 2 are more likely due to limited impact and not due to the high uncertainty such as in Study 1. Future studies are needed to determine the reason for the changes in this pattern of the causal effects.

All things considered, based on Study 1 and Study 2, it is indicative that moving around people's views regarding police behaviour influences their perceptions of the appropriateness of the law and the police through boosting their personal sense of power. Because the results regarding duty to obey the police are contradictory – and relatively sensitive – it is difficult to draw conclusions regarding that component of legitimacy. The fact that all the other alternative mediators (even the newly added self-control) remained highly sensitive and consistently non-significant gives further credibility to the primacy of sense of power.

Finally, it is worth noting that – apart from the procedurally unjust-breach of boundaries comparison – the NDEs were consistently significant, providing further support that in the absence of a strong mediating variable such as sense of power, procedural justice has a strong influence on legitimacy of the police and the law.

Study 3

Because in both Study 1 and Study 2 personal sense of power emerged as the only significant and relatively robust mediator, Study 3 was conducted to examine this particular mediator using a design-based strategy. In essence, both the parallel and

parallel encouragement designs require the same experimental procedure, they only differ in their assumptions regarding the effectiveness of the second manipulation. Accordingly, there are four different ways to derive indirect effects, which will all be tested here: parallel design with (1) no interaction assumption for each individual (point-identified), (2) no interaction assumption on average (point-identified), (3) permitting the presence of interaction (only sharp bounds), and (4) parallel encouragement design (only sharp bounds).

Participants and procedure

Study 3 was fielded on Amazon Turk, its recruitment process and the design of the first experimental manipulation was the same as for Study 1. The only slight change made was that the IP protection was extended, so Turkers who took part in Study 1 would not be allowed to join this study. In accordance with the parallel (encouragement) design, after receiving the first manipulation, half of the participants were randomly allocated to receive the second experimental manipulation and either the high sense of power or low sense of power condition. This second manipulation instructed the participants to imagine that they had been selected to a police oversight committee, which possessed a different purview depending on the manipulation. In the high power condition the committee had substantial influence, members could make decisions regarding the dismissal/promotion of police officers, and could potentially demand a major investigation into the police force. By contrast, the low power condition described a committee with a very limited purview where all decisions needed to be approved by the police chief and only recommendations could be made (for the text of the manipulation please refer to Appendix/C).

The participants were asked to reflect on their imaginary position by writing at least 300 characters (a little longer than the maximum length of a tweet). Most of the participants (n=243 or 88.7%) followed the instructions, however, 25 of them reached the limit by "cheating" (e.g., holding down one letter on their keyboard, copy-pasting meow-meow over and over again, etc.) and a further 6 decided to write about something unrelated to the task (e.g., their breakfast, their expectations regarding the next Star Wars movie's plot, etc.). Since none of these people failed on either of the other checks (i.e., attention and instructional manipulation checks) their results were left in the experiment and used for further analysis regardless.

The same pre-treatment covariates were recorded as in Study 1 and no participant decided to withdraw from the study. Altogether 570 people took part in the study. 296 of them only received the first experimental manipulation, and from those who received the second one as well, 136 were randomly assigned to the low power, 138 to the high power condition. Roughly the same number of participants read the procedurally just (n=192), procedurally unjust (n=188), and breach of boundaries (n=190) article. 51% of the participants were women (n=290), the average age was 37.7 and 52% of the participants had at least a college degree (n=298). Three-fourths of the participants were White (n=298) and they were on average slightly left-leaning (M=3.4, on a 1-7 left-right scale). 32% of them were contacted by the police (n=180) and 26% had reached out to the police in the last two years (n=148). The distribution of the pre-treatment covariates was approximately balanced across the different experimental conditions, detailed information regarding them can be found in Appendix/C. Appendix/C also contains the necessary manipulation checks for both experiments, which all showed the expected effects.

Measurements

From the scales used in Study 1, normative alignment with the police, normative alignment with the law, personal sense of power, procedural justice, and respect for boundaries were included in Study 3. Applying the same methods as in Study 1 and Study 2, the assessment of the measurement models indicated a good model fit and high internal consistency across all latent variables (for details please refer to Appendix/C).

From the four design-based approaches discussed in the methods section, only the parallel design with the no-interaction assumption in the expectation permits continuous mediators and outcomes in the current implementation of the mediation package (Tingley et al. 2014), while the remaining three methods only estimate indirect effects in the case of binary mediators and outcomes. Accordingly, normative alignment with the police and the law and sense of power were all recoded to binary variables, using their medians[6]. In particular, each variable that had a higher score than its median was given the value 1, and the lower or equal values were recoded to 0.

[6] As very similar results emerged taking the mean instead of the median, only the results for one of these transformations are shown. The results for the alternative recoding are available from the author upon request.

Results

For the design-based approaches, including pre- or post-treatment confounders are unnecessary to estimate the unbiased indirect effects of either the parallel or parallel encouragement designs thanks to the second manipulation. Table 3 includes separate estimates depending on the assumptions pursued. The first row includes the results when no interaction is assumed for each unit, the second row when no interaction is assumed on average (with the corresponding sensitivity analysis in the third row). The fourth and fifth row present the sharp bounds for the control and treatment groups when interactions are permitted, and finally, the fifth and sixth rows contain the results of the indirect effects of the compliers only among the controls and the treated.

To demonstrate the interpretation of the results, let's take the first column with the procedurally just-procedurally unjust comparison and normative alignment with the police. In the case of parallel design, when no-interaction is assumed for each individual, the natural indirect effects are significant (NIE_{noint_pjunpj}=0.408, $p<0.01$). Similarly, when the no-interaction assumption is made in the expectation, there is a relatively strong significant NIE (NIE_{effhom_pjunpj}=0.670, $p<0.01$) which is moderately sensitive to interaction heterogeneity, with the 95% confidence intervals becoming zero before reaching the average NIE ($NIE_{effhom\_pjunpj}(\sigma)$=0.891, $NIE_{effhom\_pjunpj}(\sigma_{95\%})$=0.428). When the no-interaction assumption is dropped, the NIE for the control group (i.e., procedurally unjust) does not include zero (NIE_{int_pjunpj}(control)=0.033-0.292), but it does include zero for the treatment (i.e., procedurally just) group (NIE_{int_pjunpj}(treated)=-0.240-0.238). Turning to the parallel encouragement design, among the compliers both the control and the treatment group's CNIEs are positive ($CNIE_{\_pjunpj}$(control)=0.071-0.523, $CNIE_{\_pjunpj}$(treated)=0.227-0.717), with higher values among those compliers who received the procedural justice treatment.

All in all, only the comparison between the procedurally just-procedurally unjust and procedurally just-breach of boundaries conditions yielded significant results. Among these results, when the parallel design was assumed with either of the no-interaction assumptions, the NIEs for normative alignment with the police and normative alignment with the law were all significant with higher sensitivity to interaction heterogeneity for the latter. When interactions were allowed, however, the sharp bounds rarely stayed in the positive territory, and if they did, only for the control group. By contrast, in the case of the parallel encouragement design, the CNIEs always

Personal sense of power	*Procedurally just vs procedurally unjust*		*Procedurally just vs breach of boundaries*		*Procedurally unjust vs breach of boundaries*	
	Normative alignment with the police	*Normative alignment with the law*	*Normative alignment with the police*	*Normative alignment with the law*	*Normative alignment with the police*	*Normative alignment with the law*
NIE_{noint}	0.408** [0.210, 0.618]	0.364** [0.156, 0.568]	0.398** [0.188, 0.588]	0.327** [0.135, 0.524]	-0.050 [-0.246, 0.137]	-0.029 [-0.232, 0.187]
NIE_{effhom}	0.670** [0.240, 1.098]	0.489* [0.044, 0.935]	0.563** [0.169, 0.962]	0.531** [0.125, 0.938]	-0.109 [-0.488, 0.266]	0.066 [-0.289, 0.423]
NIE_{effhom} *(σ)*	0.891 [0.428]	0.657 [0.186]	0.708 [0.294]	0.668 [0.250]	0.166 [0.000]	0.121 [0.000]
NIE_{int} *(control)*	[0.033, 0.292]	[-0.132, 0.449]	[0.049, 0.347]	[0.011, 0.514]	[0.049, 0.347]	[0.011, 0.514]
NIE_{int} *(treated)*	[-0.216, 0.238]	[-0.237, 0.218]	[-0.216, 0.238]	[-0.237, 0.218]	[0.033, 0.292]	[-0.132, 0.449]
CNIE (control)	[0.071, 0.523]	[0.033, 0.478]	[0.103, 0.689]	[-0.088, 0.587]	[0.054, 0.604]	[-0.034, 0.718]
CNIE (treated)	[0.227, 0.717]	[0.282, 0.802]	[0.274, 0.854]	[0.223, 0.753]	[0.023, 0.485]	[0.063, 0.496]

**p<0.05, **<0.01*

Table 3 Parallel and parallel encouragement design results – Study 3

stayed positive for normative alignment with the police, and among the treated for normative alignment with the law, which had consistently greater results than the control group.

Discussion

As with Study 2, the comparison of the procedurally unjust-breach of boundaries conditions did not provide significant results. For the remaining two comparisons, however, the results varied depending on which estimation strategy was pursued. Assuming that the second manipulation was perfect, and one of the two no-interaction assumptions holds true, the results for the parallel design seem to reinforce the findings from Study 1 and Study 2, and support sense of power's mediating role. In a similar vein, Study 1, Study 2, and the sensitivity analysis in the current study, all implied that the indirect effects of sense of power are relatively robust to interaction heterogeneity, which grants certain credibility to either of the no-interaction assumptions. Hence, the indirect effects that allow the presence of interactions are probably not necessary, although they emphasise how much the results are dependent on the no-interaction assumption.

A bigger concern is whether the second manipulation can be considered perfect. As noted by Imai et al. (2013), perfect manipulation of psychological constructs, such as sense of power, is inherently difficult. In addition, and as described earlier, a little over 11% of the people who received the second manipulation did not follow the instructions to the letter (i.e., they solved the task by cheating or discussed an unrelated topic), which conceivably implies some imperfection. Therefore, it is plausible that the complier natural indirect effects of the parallel encouragement design are closer to reality than the estimates of the parallel design. The CNIEs[7] never included zero for the treated and were always consistently higher for the procedurally just condition compared to the procedurally unjust or breach of boundaries conditions. Thus these results provide further support to sense of power's mediating relationship towards normative alignment of the police and normative alignment with the law.

With all these considerations in mind, the evidence base from Study 1, Study 2, and Study 3 provide a strong indication that sense of power mediates the impact of

[7] The slight differences of the CNIEs across the different comparisons is a product of the high uncertainty of the estimation which remains problematic even after increasing the number of bootstraps to a higher value.

procedural justice on normative alignment with the police, and – with qualified support – towards normative alignment with the law.

Conclusion

The results suggest that, by treating citizens in a procedurally fair way and respecting their boundaries, the police can instil a heightened sense of interpersonal control, autonomy, and power in the expectation of potential future encounters. This enhanced sense of interpersonal power seems to transmit the impact of procedural justice and to increase police and legal legitimacy in terms of the normative appropriateness of those two institutions. By comparison, the experience of strengthened mastery is less likely to boost consent to police actions and appears to be unrelated to deference to the law. It is conceivable that sense of power only mediates the impact of procedural justice on normative alignment and right to rule judgments because they are the proactive components of legitimacy, often linked to willingness to cooperate with the authorities and community engagement, whilst duty to obey is deferential, reactive, and usually associated with compliance with the law. This implies that the findings speak to the importance of policing by consent over command-and-control style policing (Jackson 2018; Tyler et al. 2015; Tyler and Jackson 2014).

While only speculation, there are two – possibly complementary – explanations as to why personal sense of power mediates the impact of procedural justice on normative alignment with the law and the police. From an instrumental point of view, it may be that subjective power brings with it a sense of increased control over desirable outcomes, with people legitimating the police in part out of self-interest. This would be in line with Thibaut and Walker's (1975) foundational work, where they found that people who were treated in a procedurally just way by the courts felt that they had more influence over the desired outcome. Alternatively, and from a normative point of view, it may be that people expect the police not to treat them as objects of suspicion and control, but rather to work with them to secure safety and maintain social order. A sense of approachability, responsiveness, and autonomy may be part of that. This would be more in line with the relational considerations put forward by Tyler and Lind (1992), and would imply that procedurally just treatment influences people's views through elevating the individuals' perceived power position, thus affirming an interpersonal notion of *policing by consent*. Future studies should examine which of these two perspectives has a better explanatory power.

The results regarding the police's grip on power were mixed. As hypothesised, it did not mediate the impact of procedural justice on either aspect of legitimacy. By contrast, there were inconsistent findings regarding the various treatments' effects, implying that if anything, procedural justice boosts the perception of the police's grip on power over individuals and their community. This implies that procedural justice not only empowers citizens, but it might also encourage them to assign more power to the authorities, potentially as a by-product of increased legitimacy. These findings are likely to dispel many of the concerns expressed by police officers (Tyler et al. 2015), as they suggest that being viewed as fair either does not have an effect or, it helps the police to be viewed as having increased power in the community.

The inclusion of self-control in Study 2 did not change the results much, and it was unaffected by procedural justice. Although there appeared to be a weak positive significant correlation between self-control and sense of power (Appendix/B), the lack of a strong relationship is not surprising, provided that sense of power is always relational (here related to the police), whilst self-control is more general. Future studies of a similar ilk might not need to include self-control as a post-treatment confounder, it might be sufficient to control for it as a dispositional, pre-treatment one.

Probably the most surprising finding of the paper is the lack of a relationship between social identification and the treatment in Study 1 and Study 2. Social identification did not mediate the various treatments' effects, and even more startlingly, the treatment only predicted social identity in one of the comparisons in Study 1. The results from Study 1 and Study 2 all seem to dispute the claims of the group-engagement model, as there is no evidence that identifying with the superordinate group would carry the impact of procedural justice to either the legitimacy of the police or the law. What could account for these unexpected findings?

One possibility is that social identification does not mediate the impact of procedural justice as suggested by the group-engagement model, rather it moderates it. This hypothesis is put forward by the group-value model (Lind and Tyler 1988; Tyler 1989), which argues that the effects of procedural justice might be dependent on one's ethnic minority background or immigration status, which could grant them a particular experience with the police that is unfamiliar to the majority of the population. As an alternative, it is also possible that procedural justice might be dependent on different levels of social identification, and people with weak/strong identification assign different importance to procedurally just cues. There are various

studies, mostly from Australia, which found mixed support for these propositions (Bradford et al. 2014; Murphy et al. 2017; Murphy and Mazerolle 2018; Murphy, Sargeant, and Cherney 2015).

As a test of the veracity of the group-value model, I examined (1) whether ethnic minority background moderated the impact of the treatment on social identity and (2) whether social identification moderated the impact of the treatment on either aspect of legitimacy. None of the fitted interactions were significant which raises doubts regarding this alternative explanation (for the detailed results from these analyses please refer to Appendix/D). Notably, social identification's main effects on the various aspects of legitimacy were often significant, and for the duty to obey constructs often stronger than the treatments' effects. Yet, if the relationship with the treatment is absent, no causal properties can be attributed to such effects. In other words, these results suggest that the relationship between social identification and legitimacy is not causal, but a mere association.

These results also advise caution regarding previous policy-relevant recommendations in the literature. Bradford (2014) for instance urged a refocus of the attention of citizenship training to procedurally just policing, arguing that it would engender inclusion. Murphy et al. (2017) called for programmes to break down barriers between the police and minority communities and asked such communities to pinpoint problematic police practices. Although both of these recommendations were intended to boost social identification by encouraging procedurally just policing, the current results suggest that the impact of procedural justice is mediated by sense of empowerment instead. This is not to say that such programmes would be useless or ineffective, but to highlight that we have a very limited understanding regarding the mechanisms (*why do they work?*) even behind very successful initiatives.

Turning towards the direct effects, manipulating views concerning appropriate police behaviour was successful in Study 1, but yielded only mixed results in Study 2 and Study 3. The comparisons between the procedurally just and the other two conditions were always significant indicating that the participants were sensitive to whether procedurally just principles were followed or were violated. However, juxtaposing the procedurally unjust and breach of boundaries conditions was more problematic, despite the manipulation checks being successful for both procedural justice and respect of boundaries (as described in Appendix/A, Appendix/B, and Appendix/C). The natural direct effects remained significant for the most part, often

even in the presence of sense of power. This shows the very strong impact of the expectations regarding appropriate police behaviour on legitimacy of the police and the law and might indicate that other psychological drivers are also responsible for transmitting its effect towards the various outcomes.

Finally, the methodological contribution of the paper is to show the versatility of statistical and design-based approaches to causal mediation analysis in tackling causal mechanisms. By making certain parametric restrictions, such as the no-interaction and the linearity assumption, causally mediated effects can be derived even in the presence of multiple mediators as shown in Study 1 and Study 2. Importantly, these assumptions are similar to those that one needs to make when relying on Structural Equation Modelling (Mackinnon 2008). Sensitivity analyses can also help to ascertain the robustness of the emerging results to the modelling assumptions and the potential for unmeasured confounding.

The design-based approach exemplified by Study 3 is clearly a more difficult endeavour. The assumption regarding the perfection of the second manipulation is often untenable in criminological research, and it is hard to imagine experiments where it could be guaranteed. Unfortunately, the parallel encouragement design does not permit point identification and usually produces noisier estimates compared to the parallel design. On balance, it is always worth estimating and comparing the results of the parallel and parallel encouragement designs, especially if the second manipulation's perfection/imperfection is in doubt. I hope that the demonstration of these methods will inspire others to follow suit, and carry out similar analysis and use similar designs in the future.

Limitations and future direction of research

One of the major limitations of the current paper is the generalisability of the results. All three of these relatively large crowdsourced convenience samples relied on highly educated participants and were younger than the population as a whole. It is quite likely that being on websites such as Amazon Turk and Prolific also comes with its own self-selection bias, which makes these samples even more peculiar. Another limitation of crowdsourcing is that the experimenter has limited control over the participants compared to a laboratory setting. This came to the front with the parallel (encouragement) design, where some people did not follow the instructions. Thus, future research should try to replicate the findings in both a more controlled

environment and by involving versatile samples that are more representative of the population.

Another potential limitation is the slight changes made for the measurement of grip on power and social identification from Study 1 to Study 2, which raises some doubts regarding the comparability of the results (even though, for social identification the change was unavoidable). Nevertheless, the comparison of the covariance matrices of Study 1 and Study 2 (Appendix/A and Appendix/B) indicated very similar patterns among the measured constructs, which makes it likely that the new measures tapped into something very similar to the old ones.

Finally, it has to be noted that causal mediation analysis relies on strong assumptions. Notably, it is often very difficult to justify the no unmeasured confounding assumption. Even after considering as many covariates as this article has, there are others which could have conceivably been included (e.g., victimisation, vicarious experience with the police, housing, employment status, etc.). Furthermore, there are other psychological processes (i.e., mediators or post-treatment confounders) which could have been missed. For instance, there is a growing literature which scrutinises how procedural justice impacts emotions (Ratcliff and Vescio 2017; Yesberg and Bradford 2018). Thus, more elaborate studies are needed to assess other potential psychological drivers of the impact of procedural justice on legitimacy of the police and the law.

Appendix/A – Measurement models, balance, and manipulation checks for Study 1

The text of the experimental manipulation

The article used for experimental manipulation read as follows (the emboldened parts varied based on the experimental condition, the first referring to the procedurally just, the second to the procedurally unjust, and the third to the breach of boundaries condition):

Stop and search practices in [State name]

On the night of 15th March around 10pm, James Williams was walking home after having watched a movie with his friends in [the second largest city in the State]. He took his phone out of his pocket to check the time when suddenly he was stopped by a couple of police officers. He was not surprised, as there was a rumor circulating about police checks in the area. He was about to put away his phone when one of them **politely asked him to keep his phone visible / yelled at him not to dare put it away, demanding he show it to them / yelled freeze and both took out their weapons, pointing them at him while demanding to show them his phone**. Mr Williams reluctantly showed them his phone holding it in front of him. One of the officers **kindly requested whether he would mind if they took a look at his phone / officers forcefully ordered him to give them the phone / forcefully took his phone out of his hand with their guns still pointed at him**. After handing over his phone, Mr Williams asked what this was all about. As a response, one of the officers **calmly explained that a smartphone similar to his had been stolen a few blocks away / one of the officers angrily shouted, commanding him to shut his mouth / one of the officers immediately forced his arms behind his back and handcuffed him, with the other officer's gun still pointing at him**. The other officer then called in on his radio that they had found a smartphone similar to the one they were looking for, citing the type and physical description of it. After a small static noise, a voice responded that this is not the phone that they were after. Subsequently the other police officer **smiled reassuringly and the one who received the news thanked Mr Williams for his cooperation and apologized for the inconvenience they might have caused / sighed impatiently and the one who received the news stared at Mr Williams suspiciously / cursed and put away his gun and the one who received the**

news uncuffed Mr Williams and threatened him not to tell anyone about the incident, finally giving back the phone to him. Following the officers departure, Mr Williams strolled home and after some hesitation decided to write a blogpost about the events. “I’m still not completely over the experience” told Mr Williams to our paper. “I’m still a bit stunned by this **fair / unfair / illegal** treatment by the police.”

In line with James Williams’ story, figures recently released by the FBI indicated that police behavior during stop and searches in [State name] was, most of the time, **professional / unprofessional / unlawful**. The number of civilian complaints have sharply **decreased from 175 in 2014 to a historically low figure of 105 / increased from 175 in 2014 to a historically high figure of 245 / increased from 175 in 2014 to a historically high figure of 245** in 2016 in [State name]. “We are aware of the changes” admitted the police chief of [the second largest city in the State], “that’s why we try to enroll as many police officers to the training programs as possible. I am sure that such efforts **are paying off / will pay off eventually / will pay off eventually**.”

Measurements with question wording

There were four dependent variables: normative alignment with and duty to obey the police, and normative alignment with and duty to obey the law. *Normative alignment with the police* was measured by four items on a 1-5 “Strongly disagree-Strongly agree” Likert-scale, and the prompt asked to what degree the respondents agreed with the statements: “The police generally have the same sense of right and wrong as I do”, “The police usually act in ways consistent with your own ideas about what is right and wrong”, “The police stand up for moral values that are important to people like me”, “The police can be trusted to make the right decisions”.

Duty to obey the police was also measured on a 1-5 Likert-scale with the question-specific response alternatives of “Not at all my duty-Completely my duty”. The prompt read “To what extent is it your moral duty to…” which was followed by three items: “...back the decisions made by the police because the police are legitimate authorities?”, “...back the decisions made by the police even when you disagree with them?”, and “...do what the police tell you even if you don't understand or agree with the reasons?”.

Normative alignment with the law had four items each measured on a 1-5 “Strongly disagree-Strongly agree” Likert-scale. People were asked to what extent

they agreed with the following statements: “Your own feelings about what is right and wrong usually agree with the laws that are enforced by the police and the courts in this neighborhood”, “The laws in your community are consistent with your own intuitions about what is right and just”, The laws of your criminal justice system are generally consistent with the views of the people in our community about what is right and wrong”, and “Obeying the law ultimately benefits everyone in the community”.

For *duty to obey the law* the prompt asked “And thinking about your duty towards the law in the United States, to what extent do you agree or disagree with the following?”. Three items were used: “All laws should be strictly obeyed”, “Even if you disagree with the law, you should always obey it”, and “Even if you do not understand why something is illegal, you should never break the law”. The response options for each of these was 1-5 “Strongly disagree-Strongly agree” Likert-scales.

The three mediators were personal sense of power, police grip on power, and social identification. *Personal sense of power* was captured using the modified scale of Anderson et al. (2012). There were four positive and four reversed items in two separate blocks, and respondents were asked to select how much they agreed with each of the items when it came to future interactions with the police (1-5 Likert-scale, Strongly disagree-Strongly agree). The four positive items were: “I can get them to listen to what I say”, “I can get them to do what I want”, “I think I have a great deal of power”, and “If I want to, I get to make decisions”. By contrast, the negative items read: “My wishes do not carry much weight”, “Even if I voice them, my views have little sway”, “My ideas and opinions are often ignored”, and “Even when I try, I am not able to get my way”.

Police grip on power was measured by two items one positive and one negative. Both of them were measured on a 1-4 Likert-scale each having item-specific response alternatives. The positive item asked “How much power do you think the police have over people like yourself?” with responses from “Little power” to “A great deal of power”. By contrast, the negative inquired “How often do you think people challenge the power of the police in your neighborhood?” with responses of “Almost never” to “Very often”.

Social identification’s scale had four items measured with 1-4 Likert-scales of “Not important at all-Very important”. Two blocks of questions asked how important it was for the participant to perceive herself/himself or to be perceived by others as “Being American” and “Being a law-abiding citizen”.

Finally, questions were asked regarding appropriate police behaviour to help assess the experimental manipulation. For *procedural justice* 1-4 Likert-scales of "Not at all often-Very often" were used. The following four questions were asked: "Based on what you have heard or your own experience, how often would you say the police generally treat people in your neighborhood with respect?", "About how often would you say that the police make fair, impartial decisions in the cases they deal with?", "When dealing with people in your neighborhood, how often would you say the police generally explain their decisions and actions when asked to do so?", and "Based on what you have heard or your own experience, how often would say the police try to do what is best for the people they are dealing with?".

Respect for boundaries was measured by six reversed items with 1-4 Likert-scales of "Not at all often-Very often". The question inquired "And how often (if ever) do you think the police in your neighborhood...?" with the following items: "...exceed their authority", "...abuse their power", "...act as if they are above the law", "...violate people's freedoms", "...get involved in situations that they have no right to be in", and "...harass and intimidate people".

Confirmatory factor analysis and internal consistency

Based on the model fit estimates, the measurement model represented the data well ($\chi^2(629)=1728.271$, $p<0.001$; RMSEA=0.066, $RMSEA_{95\%}$=0.062, 0.070; CFI=0.914, TLI=0.904; SRMR=0.071). Table 1a contains the information regarding the factor loadings and measures of internal consistency. The factor loadings were all relatively strong, with the measures of normative alignment, obligation to obey the law, appropriate police behaviour, and sense of power having higher loadings (λ=0.668-0.913) compared to the duty to obey the police, police grip on power, and social identification. (λ=0.526-0.889). Cronbach's Alphas were all above 0.8, except for police grip on power, which had a value of 0.641, partly due to being measured by only two items. The average inter-item correlation for the measures of police and legal legitimacy, personal sense of power, and appropriate police behaviour, were all above or very close to 0.6. By contrast, for social identification it was only 0.505, and for police grip on power only 0.472.

The correlations between the latent variables can be found in Table 2a. Measures of police and legal legitimacy and appropriate police behaviour had always the highest correlation with each other compared to the other measures. From the three

mediators, sense of power had the highest correlation with normative alignment with the police ($r=0.552$, $p<0.01$), normative alignment with the law ($r=0.456$, $p<0.01$), obligation to obey the police ($r=0.478$, $p<0.01$), procedural justice ($r=0.873$, $p<0.01$), and respect for boundaries ($r=0.829$, $p<0.01$); social identification had the highest correlation with obligation to obey the law ($r=0.320$, $p<0.01$). Police grip on power had the weakest correlation with the legitimacy variables ($r=0.158$-0.318, $p<0.05$-0.01). The three mediators had weak but significant relationships with each other ($r=0.200$-0.319, $p<0.01$).

Manipulation checks and balance tests

To assess whether the experimental manipulation was successful, ANOVA-analysis was used with Bonferroni-correction to account for multiple comparisons. Perception of procedural justice was significantly different across the three experimental conditions ($F(400)=71.39$, $p<0.001$), with the highest score under the procedurally just condition ($M=0.388$), followed by the procedurally unjust condition ($M=0.021$), and the breach of boundaries condition ($M=-0.402$). Respect for boundaries showed a very similar picture with significantly different average values across the three conditions ($F(400)=5.32$, $p<0.001$), the procedural justice condition having the highest score ($M=0.434$), then the procedurally unjust condition ($M=0.024$), and finally the breach of boundaries condition ($M=-0.449$). These results indicate that the textual manipulation achieved the desired effect. The covariate balance also implies that the randomisation was successful. As shown in Table 3a, all covariates of interest were approximately the same for each experimental condition.

Construct	*Number of items*	*Factor loadings (standardised)*	*Cronbach's Alpha*	*Average inter-item correlation*
Normative alignment with the police	4	0.857-0.879	0.926	0.758
Normative alignment with the law	4	0.702-0.879	0.885	0.659
Obligation to obey the police	3	0.621-0.889	0.818	0.599
Obligation to obey the law	3	0.851-0.885	0.899	0.749
Personal sense of power	8	0.668-0.917	0.941	0.665
Police grip on power	2	0.632-0.730	0.641	0.472
Social identification	4	0.526-0.877	0.803	0.505
Procedural justice of the police	4	0.793-0.865	0.903	0.699
Police respect of boundaries	6	0.826-0.913	0.920	0.768

Table 1a Factor loadings from the CFA and reliability measures – Study 1

Correlations between latent variables (CFA)	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Normative alignment with the police 1								
Normative alignment with the law 2	0.841**							
Obligation to obey the police 3	0.610**	0.551**						
Obligation to obey the law 4	0.552**	0.651**	0.631**					
Personal sense of power 5	0.561**	0.456**	0.478**	0.278**				
Police grip on power 6	0.306**	0.210**	0.318**	0.158*	0.319**			
Social identification 7	0.341**	0.357**	0.347**	0.320**	0.308**	0.200**		
Procedural justice of the police 8	0.873**	0.734**	0.529**	0.467**	0.534**	0.272**	0.277**	
Police respect of boundaries 9	0.829**	0.678**	0.453**	0.439**	0.520**	0.267**	0.208**	0.867**

Table 2a Correlation analysis results for latent variables (CFA) – Study 1

		Procedurally just condition	Procedurally unjust condition	Breach of boundaries condition	Total
Gender	Male	75	71	78	224
	Female	59	64	56	179
Age (mean, SD)		36.3 [11.4]	37.2 [12.7]	34.1 [10.4]	35.9 [11.6]
Education	High school graduate, no college	19	17	20	56
	Some college, or associate degree	53	45	49	147
	College graduate or higher degree	62	73	65	200
Ethnicity	American Indian or Alaska Native	1	1	2	4
	Hawaiian or Other Pacific Islander	1	1	0	2
	Asian or Asian American	7	19	9	35
	Black or African American	10	6	13	29
	Hispanic or Latino	6	6	6	18
	White	109	102	104	315

Political orientation (1-7, left-right scale, mean, SD)		3.3 [1.7]	3.6 [1.7]	3.3 [1.7]	3.4 [1.7]
Police initiated contact	No	90	97	92	279
	Yes	44	38	42	124
Citizen initiated contact	No	106	107	98	311
	Yes	28	28	36	92
Total		134	135	134	403

Table 3a Covariate balance – Study 1

Appendix/B – Measurement models, balance, and manipulation checks for Study 2

Measurements with question wording:

Police grip on power received two additional items in the hope that they would boost the strength of the measurement model, which they did. For each question, item-specific response alternatives were specified on a 1-4 Likert-scale. The positive item of "To what extent do you think the police have unrivalled authority over people like yourself?" used "Little authority-Great deal of authority", whilst the negative question of "How difficult is it for the police to exert their authority in your neighbourhood?" used "Very easy-Very difficult".

The scale of *social identity* needed to be changed because based on the pilot study of the experiment, "Being British" and "Being a law-abiding citizen" were unrelated to each other. Thus, a new scale was introduced with four items and 1-5 "Strongly disagree-Strongly" agree Likert-scales. Two of the items were very similar to the earlier ones, but now they were formulated as statements: "I see myself as an honest, law-abiding citizen" and "It is important to me that others see me as an honest, law-abiding citizen". The two new items were the following: "Others in my community have similar values to mine" and "Being a member of my community is important to how I see myself as a person".

Finally, a scale for *self-control* was added, the items selected from Reisig et al. (2014). Three items were positive and three were negative with the leading question: "Thinking about yourself, how accurately do these statements describe you?". Accordingly, a 1-5 Likert-scale was used with the response alternatives ranging from "Very inaccurately" to "Very accurately". The three positive items were the following: "People would say that I have iron self-discipline", "I am able to work effectively toward long-term goals", and "I am good at resisting temptation", while the negative ones were: "I have a hard time breaking bad habits", "I wish I had more self-discipline", and "I often act without thinking through all the alternatives".

Confirmatory factor analysis and internal consistency

The model fit estimates for Study 2's confirmatory factor analysis also indicated that the measurement models provided a good representation of the data ($\chi^2(944)=2035.894$, $p<0.001$; RMSEA=0.060, $RMSEA_{95\%}=0.056, 0.063$; CFI=0.917, TLI=0.909; SRMR=0.067). The factor loadings were higher for the measures of police

and legal legitimacy, sense of power, and appropriate police behaviour (λ=0.720-923). By contrast, police grip on power, social identification, and self-control had relatively lower, but still moderately strong or strong loadings (λ=0.501-0.871). Cronbach's Alphas were all higher than 0.8, except for police grip on power, which still had a strong measure of 0.769. The average inter-item correlation was higher than or very close to 0.6 for police and legal legitimacy, sense of power, and appropriate police behaviour. Social identification's Alpha coefficient was 0.585, self-control's 0.526, and police grip on power's 0.455.

Unlike in Study 1, in Study 2 personal sense of power joined the measures of legitimacy and appropriate police behaviour in having similar or larger magnitude of correlation with these measures (Table 5a). From the mediators, sense of power had the highest correlation with normative alignment with the police ($r=0.727$, $p<0.01$) and normative alignment with the law ($r=0.565$, $p<0.01$), and duty to obey the police ($r=0.385$, $p<0.01$), whilst social identification had the highest correlation with obligation to obey the law ($r=0.427$, $p<0.01$), procedural justice ($r=0.699$, $p<0.01$), and respect for boundaries ($r=0.607$, $p<0.01$). Among the mediators, personal sense of power had a significant positive relationship with police grip on power ($r=0.152$, $p<0.05$) and self-control ($r=0.158$, $p<0.05$), but seemed to be unrelated to social identification ($r=0.055$, $p>0.05$). Police grip on power was also uncorrelated with social identification ($r=-0.033$, $p>0.05$) and negatively correlated with self-control ($r=-0.158$, $p<0.01$). Finally, self-control and social identification had a small positive significant correlation ($r=0.158$, $p<0.05$). The changes in the correlations of the mediators can probably be explained by the changes in their measurement. More striking is the higher correlation of sense of power with the latent variables of police and legal legitimacy and appropriate police behaviour compared to Study 1. Reassuringly however, the changes in the magnitude of correlations by-and-large follow a very similar pattern to Study 1.

Manipulation checks and balance tests

As with Study 1, ANOVA-analysis with Bonferroni correction was used to test whether the textual manipulation was successful. Procedural justice was significantly different across the experimental conditions ($F(320)=47.59$, $p<0.001$), with the procedurally just condition having the highest average score ($M=0.524$), the procedurally unjust condition the second highest ($M=-0.072$), and the breach of

boundaries condition the lowest (M=-0.444). Respect for boundaries took on a very similar pattern ($F(320)=34.93$, $p<0.001$), the procedurally just condition having the highest mean score (M=0.447), followed by the procedurally unjust (M=-0.032) and breach of boundaries (M=-0.407) conditions. Both tests indicated that the manipulation was successful. The balance table of Table 6a also speaks to the success of the randomisation, as all covariates of interest were on average the same across the three experimental conditions.

Construct	*Number of items*	*Factor loadings*	*Cronbach's Alpha*	*Average inter-item correlation*
Normative alignment with the police	4	0.908-0.921	0.955	0.840
Normative alignment with the law	4	0.865-0.921	0.939	0.793
Obligation to obey the police	3	0.720-0.839	0.813	0.592
Obligation to obey the law	3	0.836-0.916	0.899	0.748
Personal sense of power	8	0.735-0.898	0.942	0.672
Police grip on power	4	0.501-0.870	0.769	0.455
Social identification	4	0.698-0.803	0.849	0.585
Self-control	6	0.624-0.871	0.869	0.526
Procedural justice of the police	4	0.835-0.883	0.922	0.747
Police respect of boundaries	6	0.873-0.923	0.960	0.799

Table 4a Factor loadings from the CFA and reliability measures – Study 2

Correlations between latent variables (CFA)	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Normative alignment with the police 1									
Normative alignment with the law 2	0.890**								
Obligation to obey the police 3	0.571**	0.538**							
Obligation to obey the law 4	0.466**	0.566**	0.677**						
Personal sense of power 5	0.727**	0.565**	0.385**	0.163**					
Police grip on power 6	-0.114	-0.141*	0.007	-0.234**	0.152*				
Social identification 7	0.185**	0.227**	0.377**	0.427**	0.055	-0.033			
Self-control 8	-0.018	0.004	0.127*	0.227**	0.158**	-0.084	0.158*		
Procedural justice of the police 9	0.874**	0.752**	0.436**	0.325**	0.699**	-0.122*	0.124*	-0.046	
Police respect of boundaries 10	0.797**	0.705**	0.411**	0.337**	0.604**	-0.146*	0.113	-0.037	0.913**

Table 5a Correlation analysis results for latent variables (CFA) – Study 2

		Procedurally just condition	Procedurally unjust condition	Breach of boundaries condition	Total
Gender	Male	32	39	32	103
	Female	75	68	77	220
Age (mean, SD)		35.3 [10.9]	36.3 [10.7]	35 [10.1]	35.5 [10.6]
Education	Lower levels of education	40	44	52	136
	BA degree	41	40	32	113
	Postgraduate degree	26	23	25	74
Ethnicity	Asian	9	9	10	28
	Black	2	1	1	4
	Mixed	3	2	3	88
	White	92	94	94	280
	Other	1	1	1	3
Political orientation (1-7, left-right scale, mean, SD)		3.2 [1.9]	3.0 [2.2]	2.9 [2.1]	3.1 [2.1]
Views about the Brexit decision (1-7, wrong-right, mean, SD)		3.5 [1.4]	3.6 [1.5]	3.4 [1.3]	3.5 [1.4]

Political orientation (component score, mean, SD)		0.1 [1.1]	0.0 [1.3]	-0.1 [1.2]	0.0 [1.2]
Police initiated contact	No	70	82	75	227
	Yes	37	25	34	96
Citizen initiated contact	No	64	77	80	221
	Yes	43	30	29	102
Total		107	107	109	323

Table 6a Covariate balance – Study 2

Appendix/C – Measurement models, balance, and manipulation checks for Study 3

The text of the experimental manipulation

The text of the second manipulation read as follows (the emboldened parts were different depending on the experimental condition, with the first belonging to the high power and the second to the low power condition):

As part of a state-wide police reform initiative, civilian oversight committees will be set up in [state name]. As with jurors, people from each community will be randomly invited to serve on these committees, but their identity and contribution will be kept secret from police officers. Please, imagine a situation where you were invited to one of these committees.

As a committee member, you would be tasked with evaluating the work done by the police, including their uses of force (e.g. police shootings), their patrolling activities (e.g. stop and search practices), their investigating techniques (e.g. interrogation methods), and how they carry out other police tasks in general. As a committee member, you would have **substantial / limited** influence on police work **- / , and report directly to the police chief**. Based on the information presented to you, you could **demand / propose** changes to police practices, **promote or suspend officers, and even dismiss them in case of multiple complaints / but all your recommendations would need to be approved by the police chief before being implemented**. **In addition, you would / you would not be allowed to assess the work of individual police officers, nor** have the authority to examine whether the police have followed the previous committee's **prescriptions**, **and the power to force a prosecutorial review if they have not / recommendations**. Should your work reveal serious malfeasance, **you could request a judicial review and potentially prompt a major investigation into the police force / you would need to defer to the chief of police to determine the optimal course of action**.

Please try to imagine being on this committee in such a role, and describe in a couple of sentences how you would feel.

Confirmatory factor analysis and internal consistency

The model fit indices of the confirmatory factor analysis of Study 3 also showed that the models represented the data well ($\chi^2(289)=1180.798$, $p<0.001$;

RMSEA=0.074, $RMSEA_{95\%}$=0.069, 0.078; CFI=0.941, TLI=0.933; SRMR=0.043). The factor loadings were very strong, and comparable to the values found for the same measures in Study 1 and Study 2 (λ=0.712-0.919). In a similar vein, all Cronbach Alphas were higher or close to 0.9 and the average inter-item correlation was higher than 0.65 for all variables (see Table 7a).

The correlations between personal sense of power and the other variables were relatively high (all higher than 0.5). The magnitude of the correlations with the normative alignment variables fell somewhere between Study 1 and Study 2, but the correlations between the variables for appropriate police behaviour were closer to Study 1 (Table 8a).

Manipulation checks and balance tests

As with Study 1 and Study 2, firstly the first experiment's impact was tested using ANOVAs with Bonferroni-correction. Procedural justice followed the expected distribution and was significantly different across the three conditions ($F(567)=62.99$, $p<0.001$) with the procedurally just condition having the highest average score (M=0.368) then the procedurally unjust (M=-0.033) and the breach of boundaries condition (M=-0.339). Respect of boundaries was also significantly different across the three conditions ($F(567)=43.82$, $p<0.001$) and had the very same ordering (procedurally just: M=0.380, procedurally unjust: M=-0.029, breach of boundaries: M=-0.355).

The expectation for the second experiment was that it would not influence either procedural justice or respect for boundaries, which would have indicated reverse causation. In line with these expectations, neither the means of procedural justice ($F(567)=2.86$, $p>0.05$, no 2nd experiment: M=-0.020, low power: M=-0.072, high power: M=0.115) nor the means of respect for boundaries ($F(567)=0.64$, $p>0.05$, no 2nd experiment: M=-0.015, low power: M=-0.036, high power: M=0.068) were on average different across the second experiment's conditions. However, sense of power was significantly different ($F(567)=10.52$, $p<0.001$), with the high power condition taking on a significantly higher average value (M=0.290) than the other two conditions (no 2nd experiment: M=-0.113, low power: M=-0.047). Because some of the estimations of the parallel designs will rely on binary mediators, it is crucial to assess whether the recoding of the factor scores had any impact on the experimental manipulation. Taking the recoding based on the mean of sense of power, people in the

high power condition still took on significantly higher average values ($F(567)=7.25$, $p<0.001$, high power: $M=0.645$) than the other two conditions (no 2nd experiment: $M=0.459$, low power: $M=0.463$). Thus, these results indicate that both experimental manipulations achieved the intended effect.

Finally, the covariance balance should be also assessed as a check of successful randomisation. As shown by Table 9a, all the covariate characteristics were approximately the same across the nine experimental conditions.

Construct	*Number of items*	*Factor loadings*	*Cronbach's Alpha*	*Average inter-item correlation*
Normative alignment with the police	4	0.869-0.919	0.941	0.800
Normative alignment with the law	4	0.712-0.875	0.893	0.676
Personal sense of power	8	0.745-0.909	0.946	0.684
Procedural justice of the police	4	0.822-0.879	0.915	0.729
Police respect of boundaries	6	0.833-0.909	0.955	0.777

Table 7a Factor loadings from the CFA and reliability measures – Study 3

Correlations between latent variables (CFA)	*1*	*2*	*3*	*4*
Normative alignment with the police 1				
Normative alignment with the law 2	0.834**			
Personal sense of power 3	0.668**	0.578**		
Procedural justice of the police 4	0.859**	0.816**	0.594**	
Police respect of boundaries 5	0.752**	0.672**	0.534**	0.829**

Table 8a Correlation analysis results for latent variables (CFA) – Study 3

		No second experiment			Reduced sense of power			Heightened sense of power			Total
		PJ cond.	UNPJ cond.	BoB cond.	PJ cond.	UNPJ cond.	BoB cond.	PJ cond.	UNPJ cond.	BoB cond.	
Gender	Male	52	57	45	26	19	24	19	18	20	280
	Female	49	40	53	19	26	22	27	28	26	290
Age (mean, SD)		37 [10.9]	36.3 [11.9]	37.6 [12.8]	36.7 [12.4]	39.6 [11.4]	36.7 [10.9]	38.5 [11.1]	40.7 [13.6]	39.1 [11.3]	37.7 [11.9]
Education	High school graduate, no college	8	10	12	7	5	5	4	7	1	59
	Some college, or associate degree	47	36	32	8	16	24	17	16	17	213
	College graduate or higher degree	46	51	54	30	24	17	25	23	28	298
Ethnicity	American Indian or Alaska Native	1	1	0	0	1	1	0	1	0	5
	Hawaiian or Other Pacific Islander	0	1	2	1	0	0	1	0	0	5
	Asian or Asian American	11	6	5	4	3	4	3	2	4	42
	Black or African American	7	9	13	6	5	3	1	7	2	53
	Hispanic or Latino	7	4	2	2	4	4	6	3	4	36
	White	75	76	76	32	32	34	35	33	36	429

Political orientation (1-7, left-right scale, mean, SD)		3.6 [1.6]	3.7 [1.9]	3.3 [1.6]	3.4 [1.5]	3.1 [1.5]	3.5 [1.5]	3.3 [1.6]	3.4 [1.7]	3.1 [1.8]	3.4 [1.7]
Police initiated	No	68	74	67	33	26	30	32	31	29	390
contact	Yes	33	23	31	12	19	16	14	15	17	180
Citizen initiated	No	74	80	75	36	29	31	29	33	35	422
contact	Yes	27	17	23	9	16	15	17	13	11	148
Total		101	97	98	45	45	46	46	46	46	570

Table 9a Covariate balance – Study 3

Appendix/D – Testing the propositions of the group-value model

Testing ethnic minority background's moderating role

I ran a linear regression analysis with 1000 bootstraps including all pre-treatment covariates in the models of social identification. To directly test whether ethnicity moderated the treatment's effect on the outcome, I specified an interaction between the two. As shown by the truncated Table 10a, none of the interactions were significant either in Study 1 or Study 2. The treatment significantly affected social identification only once (β_{pjbob}=0.259, $CI_{95\%}$=[0.089, 0.428], p<0.01), as already discussed with the other average treatment effects for Study 1. Ethnicity only had a significant partial association with social identity once ($\beta_{unpjbob}$=0.708, $CI_{95\%}$=[0.017, 1.400], p<0.05).

There are certain limitations why this should not be considered a full test of ethnicity's potential effect modification. First, due to the lack of ethnic diversity, ethnicity was coded as a binary variable. However, it is conceivable that specific ethnic groups might react to the treatment differently (Murphy et al. 2017). Secondly, belonging to an ethnic group does not mean that you strongly identify with it as well (Murphy et al. 2015). Overall, the measure of ethnicity is only a very crude approximation of ethnic minority background's potential effects, and more fine-grained analysis is necessary for a fuller test of it.

Testing social identification's moderating role

The potential moderating role of social identification was tested the same way as ethnic minority's was. Table 11a shows that as with the previous case, the specified interactions did not have a significant effect on either of the legitimacy variables in any of the studies. The treatments had a consistent effect on the normative alignment constructs, while social identification was associated with the duty to obey constructs in both Study 1 and Study 2; however, the treatments' and social identification's significance varied when modelling the other respective aspects of legitimacy. Ethnicity's association with the legitimacy outcomes was sporadic, and mostly emerged for the normative alignment constructs.

These models provide further evidence against the group-value model (Lind and Tyler 1988; Tyler 1989). Even when treating social identification as a dispositional characteristic, the treatment's effect does not seem to be dependent on it. However, social identification still appears to be associated with the various constructs of legal and police legitimacy, especially with their duty to obey aspect.

Linear regression analysis	*Social identification (Study 1)*			*Social identification (Study 2)*		
	Pj vs unpj	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*
Treatment	0.079	0.259**	0.187	0.007	0.039	0.002
	[-0.107, 0,266]	[0.089, 0.428]	[-0.007, 0.373]	[-0.136, 0.150]	[-0.114, 0.192]	[-0.144, 0.147]
Ethnicity	0.088	0.597	0.708*	0.220	0.179	0.198
	[-0.550, 0.726]	[-0.134, 1.328]	[0.017, 1.400]	[-0.053, 0.492]	[-0.067, 0.425]	[-0.074, 0.469]
*Treatment*Ethnicity*	0.050	-0.194	-0.289	-0.323	-0.231	0.127
	[-0.358, 0.458]	[-0.643, 0.256]	[-0.704, 0.125]	[-0.674, 0.027]	[-0.533, 0.074]	[-0.226, 0.481]

Table 10a Linear regression analysis– truncated table

Linear regression analysis		*Normative alignment with the police*			*Obligation to obey the police*			*Normative alignment with the law*			*Obligation to obey the law*		
		Pj vs unpj	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*	*Pj vs unpj*	*Pj vs bob*	*Unpj vs bob*
Study 1	*Social identity*	0.737** [0.331, 1.144]	0.500* [0.043, 0.807]	0.244 [-0.228, 0.715]	0.677** [0.152, 1.203]	0.664** [0.345, 0.917]]	0.560* [0.079, 1.041]	0.654** [0.287, 1.020]	0.605** [0.215, 0.934]	0.389 [-0.082, 0.860]	0.663** [0.195, 1.130]	0.681** [0.518, 1.102]	0.428 [-0.032, 0.887]
	Treatment	0.604** [0.431, 0.777]	1.039** [1.037, 1.489]	0.439** [0.233, 0.645]	0.563** [0.337, 0.790]	0.660** [0.445, 0.824]	0.091 [-0.133, 0.314]	0.413** [0.252, 0.573]	0.801** [0.842, 1.321]	0.399** [0.211, 0.587]	0.365** [0.180, 0.550]	0.512** [0.093, 0.521]	0.123 [-0.071, 0.317]
	Ethnicity	-0.129 [-0.330, 0.087]	-0.102** [-0.869, -0.158]	-0.050 [-0.286, 0.186]	-0.202 [-0.503, 0.098]	-0.241 [-0.265, 0.304]	-0.118 [-0.394, 0.158]	-0.019 [-0.200, 0.161]	-0.151** [-0.845, -0.136]	0.029 [-0.199, 0.257]	-0.119 [-0.340, 0.103]	-0.113 [-0.244, 0.371]	0.179 [-0.042, 0.400]
	*Social identity *Treatment*	-0.244 [-0.496, 0.008]	-0.111 [-0.373, 0.596]	0.149 [-0.142, 0.439]	-0.108, [-0.466, 0.250]	-0.110, [-0.277, 0.559]	0.003 [-0.304, 0.310]	-0.192 [-0.417, 0.034]	-0.165 [-0.521, 0.369]	0.047 [-0.227, 0.320]	-0.200, [-0.492, 0.091]	-0.229 [-0.463, 0.384]	0.004 [-0.290, 0.298]

Study 2	*Social identity*	0.407* [0.062, 0.752]	0.425* [0.054, 0.796]	0.372 [-0.001, 0.745]	0.581** [0.293 ,0.869]	0.631** [0.336, 0.926]	0.648** [0.368, 0.928]	0.540** [0.184, 0.894]	0.574** [0.224, 0.925]	0.536** [0.169, 0.902]	0.813** [0.542, 1.084]	0.810** [0.500, 1.120]	0.828** [0.506, 1.149]
	Treatment	0.892** [0.674, 1.111]	1.263** [1.013, 1.513]	0.374** [0.102, 0.647]	0.434** [0.262, 0.606]	0.634** [0.449, 0.820]	0.183 [-0.019, 0.385]	0.704** [0.490, 0.917]	1.081** [0.850, 1.313]	0.390** [0.104, 0.675]	0.164 [-0.038, 0.365]	0.307** [0.098, 0.516]	0.130 [-0.089, 0.350]
	Ethnicity	-0.385 [-0.808, 0.039]	-0.513** [-0.863, -0.164]	-0.504* [-0.919, -0.089]	0.107 [-0.174, 0.388]	0.020 [-0.256, 0.296]	-0.200 [-0.523, 0.123]	-0.407 [-0.842, 0.029]	-0.491** [-0.865, -0.117]	-0.642** [-1.143, -0.141]	0.025 [-0.328, 0.378]	0.063 [-0.225, 0.351]	-0.262 [-0.624, 0.100]
	*Social identity *Treatment*	0.136 [-0.337, 0.609]	0.111 [-0.362, 0.585]	0.029 [-0.475, 0.532]	0.206 [-0.195, 0.606]	0.141 [-0.279, 0.560]	-0.007 [-0.411, 0.398]	-0.036 [-0.503, 0.432]	-0.076 [-0.515, 0.363]	0.003 [-0.492, 0.497]	-0.01 [-0.427, 0.398]	-0.039 [-0.471, 0.392]	0.021 [-0.401, 0.442]

Table 11a Linear regression analysis – truncated table

References

Anderson, Cameron, Oliver P. John, and Dacher Keltner. 2012. "The Personal Sense of Power." *Journal of Personality* 80(2):313–44.

Anduiza, Eva and Carol Galais. 2016. "Answering Without Reading: IMCs and Strong Satisficing in Online Surveys." *International Journal of Public Opinion Research* 29(3):1–23.

Blader, Steven L. and Tom R. Tyler. 2009. "Testing and Extending the Group Engagement Model: Linkages between Social Identity, Procedural Justice, Economic Outcomes, and Extrarole Behavior." *Journal of Applied Psychology* 94(2):445–64.

Bradford, Ben. 2014. "Policing and Social Identity: Procedural Justice, Inclusion and Cooperation between Police and Public." *Policing and Society* 24(1):22–43.

Bradford, Ben, Katrin Hohl, Jonathan Jackson, and Sarah MacQueen. 2015. "Obeying the Rules of the Road." *Journal of Contemporary Criminal Justice* 31(2):171–91.

Bradford, Ben, Jenna Milani, and Jonathan Jackson. 2017. "Identity, Legitimacy and 'Making Sense' of Police Use of Force." *Policing: An International Journal of Police Strategies & Management* 40(3):614–27.

Bradford, Ben, Kristina Murphy, and Jonathan Jackson. 2014. "Officers as Mirrors." *British Journal of Criminology* 54(4):527–50.

Bullock, John G., Donald P. Green, and Shang E. Ha. 2010. "Yes, but What's the Mechanism? (Don't Expect an Easy Answer)." *Journal of Personality and Social Psychology* 98(4):550–58.

Gan, Muping, Daniel Heller, and Serena Chen. 2018. "The Power in Being Yourself : Feeling Authentic Enhances the Sense of Power." *Personality and Social Psychology Bulletin* In Press. Retrieved (http://journals.sagepub.com/doi/abs/10.1177/0146167218771000).

Hamm, J. A., R. Trinkner, and J. D. Carr. 2017. "Fair Process, Trust, and Cooperation: Moving Toward an Integrated Framework of Police Legitimacy." *Criminal Justice and Behavior* 44(9):1183–1212.

Hauser, David J. and Norbert Schwarz. 2016. "Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants." *Behavior Research Methods* 48(1):400–407.

Hough, Mike. 2012. "Researching Trust in the Police and Trust in Justice: A UK

Perspective." *Policing and Society* 22(3):332–45.

Hough, Mike, Jonathan Jackson, and Ben Bradford. 2013. "Legitimacy, Trust and Compliance: An Empirical Test of Procedural Justice Theory Using the European Social Survey." Pp. 326–53 in *Legitimacy and Criminal Justice - An International Exploration*, edited by J. Tankebe and A. Liebling. Oxford University Press.

Huq, A. Z. Aziz H., J. Jackson, and R. J. Trinker. 2017. "Legitimating Practices: Revisiting the Predicates of Police Legitimacy." *British Journal of Criminology* (57):1101–22.

Imai, K., D. Tingley, and T. Yamamoto. 2013. "Experimental Designs for Identifying Causal Mechanisms." *Journal of the Royal Statistical Society Series A-Statistics in Society* 176(1):5–51.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105(4):765–89.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010. "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science* 25(1):51–71.

Imai, Kosuke and Teppei Yamamoto. 2013. "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments." *Political Analysis* 21(2):141–71.

Jackson, Jonathan et al. 2012. "Why Do People Comply with the Law?" *British Journal of Criminology* 52(6):1051–71.

Jackson, Jonathan. 2018. "Norms, Normativity, and the Legitimacy of Justice Institutions: International Perspectives." *Annual Reveiw of Law and Social Sciences* 14 In pres.

Jackson, Jonathan, Ben Bradford, Mike Hough, and Stephany Carrillo. 2014. *Extending Procedural Justice Theory - A Fiducia Report on the Design of New Survey Indicators*. Retrieved (http://eprints.lse.ac.uk/62237/1/Extending procedural justice theory.pdf).

Jackson, Jonathan and Jacinta M. Gau. 2015. "Carving up Concepts? - Differentiating between Trust and Legitimacy in Public Attitudes towards Legal Authority." Pp. 49–69 in *Interdiisciplinary Perspectives on Trust - Towards Theoretical and Methodological Integration*, edited by E. Shockley, T. M. S. Neal, L. PytlikZillig,

and B. Bornstein. Springer.

Jackson, Jonathan and Jason Sunshine. 2007. “Public Confidence in Policing: A Neo-Durkheimian Perspective.” *British Journal of Criminology* 47(2):214–33.

Keele, Luke. 2015. “Causal Mediation Analysis Warning! Assumptions Ahead.” *American Journal of Evaluation* 46(4):500–513.

Keele, Luke, Dustin Tingley, and Teppei Yamamoto. 2015. “Identifying Mechanisms behind Policy Interventions via Causal Mediation Analysis.” *Journal of Policy Analysis and Management* 34(4):937–63.

Light, Alysson E., Kimberly Rios, and Kenneth G. DeMarree. 2018. “Self-Uncertainty and the Influence of Alternative Goals on Self-Regulation.” *Personality and Social Psychology Bulletin* 44(1):24–36.

Lind, Allan and Tom Tyler. 1988. *The Social Psychology of Procedural Justice*. Springer.

Loader, Ian. 2006. “Policing, Recognition, and Belonging.” *Annals of the American Academy of Political and Social Science* 605(1):201–21.

Mackinnon, David P. 2008. *Introduction to Statistical Mediation*. Erlbaum.

MacQueen, Sarah and Ben Bradford. 2015. “Enhancing Public Trust and Police Legitimacy during Road Traffic Encounters: Results from a Randomised Controlled Trial in Scotland.” *Journal of Experimental Criminology* 11(3):419–43.

Maglio, Sam J., Yaacov Trope, and Nira Liberman. 2013. “The Common Currency of Psychological Distance.” *Current Directions in Psychological Science* 22(4):278–82.

Malhotra, Neil. 2008. “Completion Time and Response Order Effects in Web Surveys.” *Public Opinion Quarterly* 72(5):914–34.

Mazerolle, Lorraine, Emma Antrobus, Sarah Bennett, and Tom R. Tyler. 2013. “Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice.” *Criminology* 51(1):33–63.

Meares, Tracey. 2017. “Policing and Procedural Justice: Shaping Citizens’ Identities to Increase Democratic Participation.” *Northwestern University Law Review* 111(6):1525–35.

Mentovich, Avital. 2012. *The Power of Fair Procedures - The Effect of Procedural Justice on Perceptions of Power and Hierarchy*. New York University.

Mooijman, Marlon et al. 2017. “Resisting Temptation for the Good of the Group:

Binding Moral Values and the Moralization of Self-Control." *Journal of Personality and Social Psychology* In Press. Retrieved (http://doi.apa.org/getdoi.cfm?doi=10.1037/pspp0000149).

Moravcová, Eva. 2016. "Willingness to Cooperate with the Police in Four Central European Countries." *European Journal on Criminal Policy and Research* 22(1):171–87.

Murphy, K., B. Bradford, and J. Jackson. 2016. "Motivating Compliance Behavior Among Offenders: Procedural Justice or Deterrence?" *Criminal Justice and Behavior* 43(1):102–18.

Murphy, Kristina, Robert J. Cramer, Kevin A. Waymire, and Julie Barkworth. 2017. "Police Bias, Social Identity, and Minority Groups: A Social Psychological Understanding of Cooperation with Police." *Justice Quarterly* In press:1–26. Retrieved (http://doi.org/10.1080/07418825.2017.1357742).

Murphy, Kristina and Lorraine Mazerolle. 2018. "Policing Immigrants : Using a Randomized Control Trial of Procedural Justice Policing to Promote Trust and Cooperation." *Australian & New Zealand Journal of Criminology* 51(1):3–22.

Murphy, Kristina, Elise Sargeant, and Adrian Cherney. 2015. "The Importance of Procedural Justice and Police Performance in Shaping Intentions to Cooperate with the Police: Does Social Identity Matter?" *European Journal of Criminology* 12(6):719–38.

Oppenheimer, Daniel M., Tom Meyvis, and Nicolas Davidenko. 2009. "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power." *Journal of Experimental Social Psychology* 45(4):867–72.

Pearl, Judea. 2001. "Direct and Indirect Effects." *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* 411–20.

van Prooijen, Jan-Willem, Kees van den Bos, and Henk a M. Wilke. 2002. "Procedural Justice and Status: Status Salience as Antecedent of Procedural Fairness Effects." *Journal of Personality and Social Psychology* 83(6):1353–61.

van Prooijen, Jan Willem. 2009. "Procedural Justice as Autonomy Regulation." *Journal of Personality and Social Psychology* 96(6):1166–80.

Radburn, Matthew and Clifford Stott. 2018. "The Social Psychological Processes of ' Procedural Justice ': Concepts , Critiques and Opportunities." *Criminology and Criminal Justice* In Press. Retrieved (http://journals.sagepub.com/doi/abs/10.1177/1748895818780200).

Ratcliff, Nathaniel J. and Theresa K. Vescio. 2017. “The Effects of Leader Illegitimacy on Leaders’ and Subordinates’ Responses to Relinquishing Power Decisions.” *European Journal of Social Psychology* 48(3):365–79.

Reisig, Michael D., Justice Tankebe, and Gorazd Mesko. 2014. “Compliance with the Law in Slovenia: The Role of Procedural Justice and Police Legitimacy.” *European Journal on Criminal Policy and Research* 20(2):259–76.

Robins, James M. and Sander Greenland. 1992. “Identifiability and Exchangeability for Direct and Indirect Effects.” *Epidemiology* 3(2):143–55.

Spencer, Steven J., Mark P. Zanna, and Geoffrey T. Fong. 2005. “Establishing a Causal Chain: Why Experiments Are Often More Effective than Mediational Analyses in Examining Psychological Processes.” *Journal of Personality and Social Psychology* 89(6):845–51.

De Stavola, Bianca L., Rhian M. Daniel, George B. Ploubidis, and Nadia Micali. 2015. “Mediation Analysis with Intermediate Confounding: Structural Equation Modeling Viewed through the Causal Inference Lens.” *American Journal of Epidemiology* 181(1):64–80.

Stavola, Bianca L. De, Rhian M. Daniel, George B. Ploubidis, and Nadia Micali. 2015. “Practice of Epidemiology Mediation Analysis With Intermediate Confounding : Structural Equation Modeling Viewed Through the Causal Inference Lens.” 181(1):64–80.

Steen, Johan, Tom Loeys, Beatrijs Moerkerke, and Johan Steen. 2017. “Flexible Mediation Analysis with Multiple Mediators.” *American Journal of Epidemiology* 186(2):184–93.

Taguri, Masataka, John Featherstone, and Jing Cheng. 2018. “Causal Mediation Analysis with Multiple Causally Non-Ordered Mediators.” *Statistical Methods in Medical Research* 27(1):3–19.

Tchetgen Tchetgen, Eric J. and Tyler J. VanderWeele. 2014. “Identification of Natural Direct Effects When a Confounder of the Mediator Is Directly Affected by Exposure.” *Epidemiology*. 25(2):282–91.

Thibaut, John and Laurens Walker. 1975. *Procedural Justice: A Psychological Analysis*. Lawrence Erlbaum Associates.

Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. 2014. “Mediation: R Package for Causal Mediation Analysis.” *Journal of Statistical Software* 59(5):1–38.

Tourangeau, Roger, Mick P. Couper, and Frederick G. Conrad. 2013. “Up Means Good: The Effect of Screen Position on Evaluative Ratings in Web Surveys.” *Public Opinion Quarterly* 77(S1):69–88.

Trinkner, Rick, Jonathan Jackson, and Tom R. Tyler. 2017. “Bounded Authority: Expanding ‘Appropriate’ Police Behavior Beyond Procedural Justice.” *Law and Human Beh* 42(3):280–93.

Trinkner, Rick and Tom R. Tyler. 2016. “Legal Socialization : Coercion versus Consent in an Era of Mistrust.” *Annual Review of Law and Social Science* 12:417–39.

Tyler, Phillip Atiba Goff, and Robert J. MacCoun. 2015. “The Impact of Psychological Science on Policing in the United States: Procedural Justice, Legitimacy, and Effective Law Enforcement.” *Psychological Science in the Public Interest* 16(3):75–109.

Tyler, Tom. 1989. “The Psychology of Procedural Justice: A Test of the Group-Value Model.” *Journal of Personality and Social Psychology* 57(5):830–38.

Tyler, Tom. 2009. “New Approaches to Justice in the Light of Virtues and Problems of the Penal System.” Pp. 19–38 in *Social psychology of punishment of crime*, edited by M. E. Oswald, S. Bieneck, and J. Hupfeld-Heinemann. Wiley.

Tyler, Tom R. and Steven L. Blader. 2003. “The Group Engagement Model: Procedural Justice, Social Identity, and Cooperative Behavior.” *Personality and Social Psychology Review* 7(4):349–61.

Tyler, Tom R. and Jonathan Jackson. 2013. “Future Challenges in the Study of Legitimacy and Criminal Justice.” Pp. 83–104 in *Legitimacy and Criminal Justice - An International Exploration*, edited by J. Tankebe and A. Liebling. Wiley.

Tyler, Tom R. and Jonathan Jackson. 2014. “Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation, and Engagement.” *Psychology, Public Policy, and Law* 20(1):78–95.

Tyler, Tom R. and Allan E. Lind. 1992. “A Relational Model of Authority in Groups.” *Advances in Experimental Social Psychology* 25:115–91.

VanderWeele, Tyler J. 2016. “Mediation Analysis: A Practitioner’s Guide.” *Annual Review of Public Health* 37(1):17–32.

VanderWeele, Tyler J. and Stijn Vansteelandt. 2014. “Mediation Analysis with Multiple Mediators.” *Epidemiologic Methods* 2(1):95–115.

Vazsonyi, Alexander T., Gabriela Ksinan Jiskrova, Albert J. Ksinan, and Marek Blatný. 2016. “An Empirical Test of Self-Control Theory in Roma Adolescents.” *Journal of Criminal Justice* 44:66–76.

Vazsonyi, Alexander T., Jakub Mikuška, and Erin L. Kelley. 2017. “It’s Time: A Meta-Analysis on the Self-Control-Deviance Link.” *Journal of Criminal Justice* 48:48–63.

Yesberg, Julia and Ben Bradford. 2018. “Affect and Trust as Predictors of Public Support for Armed Police : Evidence from London.” *Policing and Society* In Press.

Interlude 2

Paper 2 identified personal sense of power as a mediator of the impact of procedural justice on one aspect of police and legal legitimacy, that is, normative alignment. The analysis took advantage of both statistical (i.e., semi-parametric structural equation modelling) and design-based (i.e., parallel (encouragement) design) approaches and arrived at very similar conclusions, if with varying certainty and effect sizes. It is encouraging that, despite the increasing methodological rigour and strong (often testable and quantifiable) assumptions, the findings were persistent.

The lack of significant findings for social identification was surprising and led me to examine alternative theoretical accounts without much success. Should other studies capable of identifying causal effects support the findings of Paper 2, it would call into question a line of research that has been based almost exclusively on observational data in policing research. If one of the key psychological mechanisms triggered by procedural justice is, indeed, empowerment, this would have strong policy relevance and indicate that existing training and citizenship programmes might emphasise more individual autonomy and mastery instead of encouraging identification with the superordinate group.

Paper 3 returns to the ScotCET dataset to continue the work on personal sense of power. Unlike for social identification, so far only a very limited number of studies has scrutinised the potential impact of sense of power on legitimacy and societally desirable outcomes, such as willingness to cooperate or compliance with the law. In particular, sequential models have been tested where both personal sense of power and procedural justice have been treated as alternative mediators with their results being juxtaposed. From the two aspects of legitimacy, Paper 3 only addresses duty to obey and differentiates between a "normative" (consensual) and "non-normative" (prudential) understanding of it.

Paper 3: "Truly Free Consent"? Clarifying the Nature of Police Legitimacy Using Causal Mediation Analysis

Krisztián Pósch, Jonathan Jackson, Ben Bradford, Sarah MacQueen

Abstract

Objectives To disentangle people's normative and non-normative forms of obligation to obey the police. To test whether normative and non-normative forms of obligation relate in diametrically opposed ways to procedural justice, personal sense of power, compliance, and cooperation. To illustrate a new approach to causal mediation analysis in the context of a randomised controlled trial.

Methods Implementation failure in the Scottish Community Engagement Trial (ScotCET) meant that this block randomised experiment designed to test procedurally just policing had a putative but unexpectedly negative causal effect. While a recent assessment indicated that it is meaningful to assess the treatment effect because of treatment consistency and homogeneity across the 20 blocks and no sign of selection bias (Pósch 2018, Paper 1 in the thesis), the unexpected direction of the treatment effect increases uncertainty about how it was transmitted. To help extract value from the study we used a natural effect model for causally ordered mediators (Steen, Moerkerke, and Vansteelandt, 2017) to assess causal pathways that include but also extend beyond treatment to procedural justice.

Results First, confirmatory factor analysis indicates that normative obligation and non-normative obligation are empirically distinct. Second, normative obligation operates as expected within a procedurally just policing framework. It responds positively to procedural justice and personal sense of power (linked back to the treatment) and mediates the treatment's influence on intentions to cooperate and comply. But non-normative obligation does not operate as expected: despite being affected by the treatment it does not carry the treatment's effect on cooperation or compliance. Overall, normative obligation to obey emerges as the most important causal mediator for cooperation, while for legal compliance normative obligation is complemented by sense of power or procedural justice.

Conclusion Criminology has yet to properly address the challenge of causal mediation analysis. We illustrate a variant of an emerging set of statistical techniques and causal identification criteria that have been developed outside of the discipline but will be of interest to readers of this journal. We also argue that duty to obey can reasonably be equated with legitimacy in the current context so long as it is properly measured.

Key words: causal mediation analysis; measurement; natural effect models; obligation to obey the police; police legitimacy; procedural justice

Introduction:

> 'Legitimacy is a psychological property of an authority, institution, or social arrangement that leads those connected to it to believe that it is appropriate, proper, and just. Because of legitimacy, people feel that they ought to defer to decisions and rules, following them voluntarily out of obligation rather than out of fear of punishment or anticipation of reward.' (Tyler 2006a: 375).

As the right to power and the authority to govern, legitimacy is central to crime-control policy. On the one hand, legitimacy reduces the tension between power-holders and subordinates (Tyler and Jackson 2013, 2014). When people view legal authorities as appropriate, proper and just, they feel a normatively grounded duty to comply with the law (Murphy, Bradford, and Jackson 2016; Murphy, Tyler, and Curtis 2009; Slocum, Ann Wiley, and Esbensen 2016; Sunshine and Tyler 2003) and cooperate with the police and criminal courts (Huq, Tyler, and Schulhofer 2011a; Huq, Tyler, and Schulhofer 2011b; Reisig and Lloyd 2008; Wolfe et al. 2016). On the other hand, legitimacy constrains power in normatively appropriate ways. To be seen as legitimate, authority figures need to treat individuals with respect and dignity, make decisions in open, neutral and accountable ways, and respect the limits of their rightful authority (Bradford, Murphy, and Jackson 2014; Jonathan-Zamir and Harpaz 2018; Murphy and Cherney 2012). On this account, legitimacy forms part of a virtuous circle. By tilting the authority-citizen relationship from coercive to consensual, legitimacy reduces the need for costly and minimally effective forms of crime-control, opening up space for

policing strategies that prioritise consent over coercion (Anon 2015; Tyler, Goff, and MacCoun 2015).

But is this portrayal of legitimacy and power relations overly optimistic? A central proposition of procedural justice theory (Tyler 2006a, 2006b) is that people feel a moral obligation to obey the rules and orders that emanate from an institution that they believe wields its power in normatively appropriate ways (i.e., has the right to power). As described in Obama's Taskforce for 21st Century Policing (2015: 5), the notion of *truly free consent* is central to the legitimacy concept: people believe that the police have the right to tell people what to do; they feel an obligation to obey because they believe that it is the right thing to do, not because they fear punishment, or feel powerless to do otherwise. To measure obligation to obey, research participants are generally asked to agree or disagree with attitudinal statements such as "you should accept police decisions because it is the right or proper thing to do" and "you should obey the orders of police officers even if you disagree with them" (Jackson 2018; Jackson and Gau 2015; Tyler and Jackson 2013).

Yet, scholars (Bottoms and Tankebe 2012; Tankebe 2009: 1279-1281; Tankebe 2013: 105-106; see also Johnson, Maguire, and Kuhns 2014: 970) have recently raised the possibility that standard measures may conflate normative and non-normative forms of obligation. In particular, research participants could report feeling an obligation to obey the police, not only because of legitimacy (*it's my freely-chosen duty as a citizen to allow officers to dictate appropriate behaviour*), but also because of pragmatism (*it's not worth risking non-compliance*) and dull compulsion (*it's not my place to question the orders of police*). It could even be that someone who experiences their relationship to the police as a 'power relationship, pure and simple, with no element of right' (Bottoms and Tankebe 2012: 126) could report feeling obligated to obey the police (Tankebe 2009: 1279-1281; Tankebe 2013: 105-106; see also Johnson et al. 2014: 970). If it is true that prior studies have failed to distinguish between normative and non-normative forms of obligation to obey, then we should reconsider the concept of legitimacy, the extant evidence base, and the policy prescription that flows from this important body of research.

In this paper, we respond to Bottoms and Tankebe's (2012) call for the disentanglement of motives in people's obligation to obey to clarify the nature of obligation to obey. As they argue (p. 165):

> "…there are several reasons other than true legitimacy why people might express feelings of obligation to obey the law: these include structurally-generated apathy and pragmatic acquiescence (dull compulsion) and instrumental calculations. To measure true legitimacy, these alternative motives need to be disentangled; however, most existing studies have not paid sufficient attention to the need for this disentanglement."

By way of contribution, we draw on data from a randomised controlled trial (RCT) of procedurally fair traffic policing (MacQueen and Bradford 2015, 2017) to address two connected questions related to the nature and measurement of police legitimacy:

1. Can normative (consensual) and non-normative (prudential) forms of obligation be teased apart empirically?
2. If they can be teased apart empirically, do they exhibit diametrically opposed dynamics in a procedural justice model of regulatory police-citizen encounters?

We address the first question by drawing on the RCT's survey data that fielded (hopefully more precise) measures of normative and non-normative forms of obligation. To assess their empirical distinctiveness, we use confirmatory factor analysis. This is an important first step, because if they cannot be disentangled, it is difficult to assess whether they respond differently to police-citizen contact. To address the second research question, we place both forms of obligation within procedural justice theory (as tested by the RCT). We assess whether the measures designed to capture normative obligation carry the effect of previous contact with the police on cooperation and compliance (in ways expected by procedural justice theory). We also evaluate the relationship between non-normative obligation and, among other things, procedural justice, sense of power, normative obligation, cooperation, and compliance.

We conclude that normative obligation can be reasonably treated as the kind of duty to obey that is consistent with Tyler's (2006a, 2006b) conceptual definition of legitimacy. But, our findings shed light on the dynamics of both consensual *and* coercive police-citizen relations (cf. Tyler et al. 2015). In particular, normative obligation appears to be sensitive to procedurally just or unjust police behaviour while

non-normative obligation seems to be rather 'sticky' and unresponsive. Indeed, non-normative obligation appears to exist outside of the procedural justice theory. It does not transmit the impact of the contact on either cooperation or legal compliance and does not have a correlation with normative obligation, however, it has moderately strong negative correlations with procedural justice and sense of power.

The paper also makes a methodological contribution in the context of causal mediation analysis. The challenge inherent in estimating the mechanism through which a causal effect is transmitted is often under-appreciated in criminology, yet developments have emerged in other disciplines for testing direct and indirect effects that go beyond to the standard *product method* associated with Baron and Kenny (1986). In this paper, we illustrate the use of a natural effect model for causally ordered mediators (Steen, Loeys, Moerkerke, and Steen 2017; Steen, Loeys, Moerkerke, and Vansteelandt 2017). This technique – which better frames the problem and more precisely estimates causal mediation effects – has not yet (to our knowledge) been applied within the discipline. In fact, we argue that the current RCT is a particularly apposite application of this methodological tool. The RCT suffered from an unusual type of implementation failure; there was a treatment effect, but it was in the opposite direction expected, and this increases uncertainty about how the causal effect was transmitted. We argue that value can be extracted, so long as sufficient methodological care is taken.

We proceed as follows. First, we consider the nature of legitimacy. Second, we expand on the significance of the research problem and detail the empirical and theoretical goals of the current study. Third, we consider the challenge of causal mediation analysis, recent statistical analyses, and the RCT's implementation failure. Fourth, we outline the study's design and findings. We close the paper with three main conclusions: (a) that normative obligation can reasonably be included in the legitimacy concept, so long as it is properly defined and measured; (b) that normative and non-normative aspects of duty to obey are unrelated, and the latter does not seem to mediate the contact's effect on cooperation and compliance; and (c) that causal mediation analysis is a flexible and effective tool which should be more widely employed in criminology when indirect effects are under scrutiny.

What is legitimacy?

Scholars typically, but not universally, think of legitimacy as having two constituent parts: (i) the right to power and (ii) the authority to govern (Bottoms and Tankebe 2012; Hamm, Trinkner, and Carr 2017; Jackson and Gau 2015; Tyler 2006a, 2006b, Tyler and Jackson 2013, 2014). The police have legitimacy in the eyes of citizens when those they serve and protect (a) view the institution as normatively appropriate (the perceived right to power) and (b) internalise the overarching moral value that they should obey orders and accept decisions because of the source not because of the content (the belief that an institution has the authority to dictate appropriate behaviour).

Right to power judgements have been operationalised in a number of different ways. One way is institutional trust (Tyler, 2006a, 2006b). To measure institutional trust, research participants are asked whether they believe that officers wield their authority in ways that take into account the interests of citizens and society – two indicative agree/disagree statements are '*the police can be trusted to make decisions that are right for your community'* and 'w*hen the police deal with people they almost always behave according to the law'* (see Sunshine and Tyler 2003; Tyler, Fagan, and Geller 2014; Tyler, Schulhofer, and Huq 2010). Normative alignment is another way of operationalising appropriateness judgements, with respondents asked whether police officers share important values and act in ways that accord with societal norms about how to exercise authority. Two example indicators are '*the police usually act in ways that are consistent with my own ideas about what is right and wrong'* and '*the police stand up for values that are important to you'* (see Jackson et al. 2012; Tyler and Jackson 2014; Tyler, Jackson, and Mentovich 2015). Both approaches measure a judgment of the normative appropriateness of the institution. Institutional trust focusses on the trustworthiness of officers to wield power in ways that take into account the interests of the public (activating a willingness to be vulnerable among citizens), while normative alignment focusses on the belief that officers act in normatively appropriate ways (activating reciprocal norms to behave appropriately as citizens).

The second part of the legitimacy concept echoes the Weberian insight that power is transformed into authority when it has popular legitimacy (Tyler 2003, 2004). The perceived entitlement to enforce the law and dictate appropriate behaviour is rooted in the institutional normativity that grants individual officers the right to dictate

appropriate behaviour in certain prescribed circumstances. As already noted, Tyler and others see this as a form of obligation that is rooted in willing authorisation and consent (Tyler and Jackson 2013). When one believes that an institution is entitled to be obeyed, one treats orders and rules as superseding one's own judgement (one obeys because of the source not the content). This is a form of deference that is connected to the rights and responsibility of legal citizenship, not because of powerlessness (a prudential "path of least resistance") or fear (of the consequences of non-compliance).

The conceptual and theoretical contribution

The current paper focusses on the second aspect of legitimacy, given the current debate about the nature of obligation. To test the open and empirical question of whether normative and non-normative forms of obligation to obey the police can be disentangled empirically, the RCT fielded measures of normative and instrumental forms of obligation that were designed to have stronger face validity than existing survey items. The measures of normative obligation were designed to elicit truly free consent (e.g., '*I feel a moral obligation to obey the police*' and '*I feel a moral duty to support the decisions of police officers, even if I disagree with them*') while the measures of non-normative obligation were designed to elicit more non-normative motives (e.g., '*People like me have no choice but to obey the police*' and '*If you don't do what the police tell you they will treat you badly')*. Our analysis has two stages:

1. We assess whether they are empirically distinct using latent variable modelling; and,
2. We estimate (assuming they are empirically distinct) whether they operate differently within an extended procedural justice framework.

To foreshadow the results of the first stage, confirmatory factor analysis indicates that they do seem to represent two distinct constructs, i.e., that normative and non-normative forms of obligation do not 'overlap' considerably and thus may not 'move around' in tandem. So, given their empirical distinctiveness, what are their dynamics in the context of police-citizen encounters? Can the first form of obligation be best characterised as normative (and as such be included in the legitimacy concept)? Does, for instance, a procedurally just encounter with the police seem to increase normative obligation? What about the dynamics of the non-normative form of

obligation? Might the experience of procedural justice decrease non-normative obligation and how does this form of obligation relate to compliance and cooperation? Figure 1 summarises the theoretical framework, where the treatment on the left-hand side represents whether the research participant was in the control group in the RCT ('business as usual') or the treatment group (the intended 'procedurally just' encounter). In Figure 1 solid lines indicate a posited influence while dashed lines indicate an uncertain predicted relationship where null or negative effects are also possible (given the lack of research on non-normative obligation).

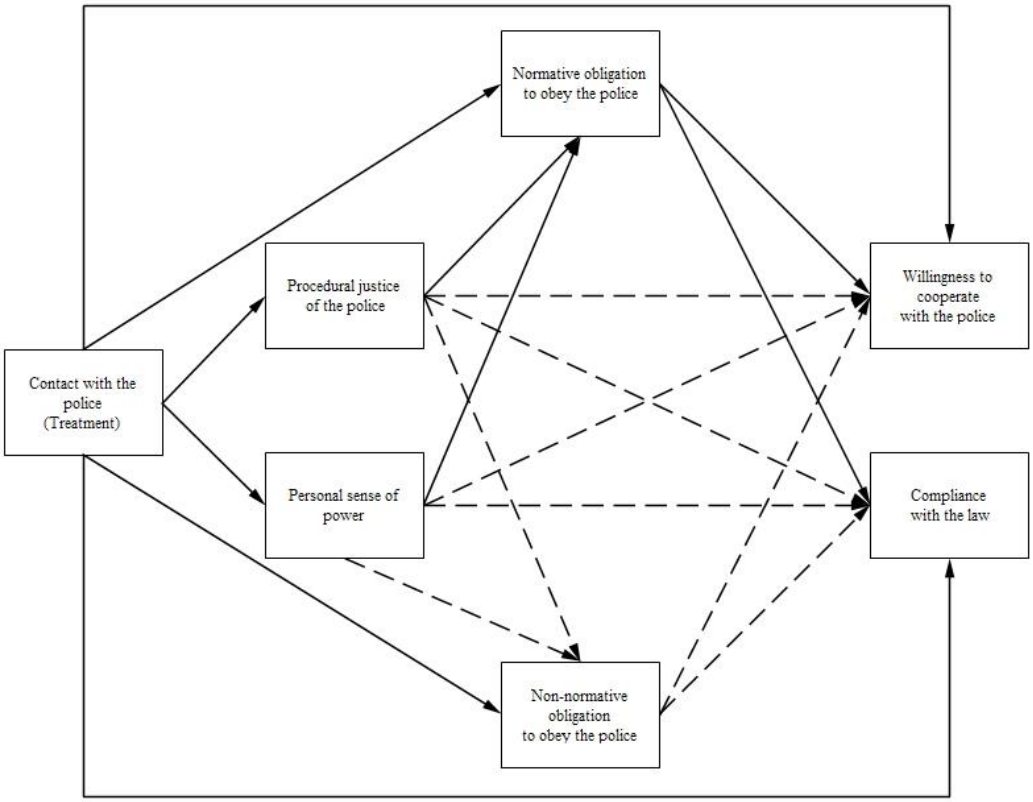


Figure 1 Theoretical model for cooperation and compliance with two pairs of sequentially ordered mediators

Normative obligation

If normative obligation is indeed normative, we would expect a particular causal chain. The first factor relates to procedural justice. We would expect the induced variation in experienced procedural justice (in the encounter with the officers) to positively predict normative obligation. Procedural fairness is a key societal norm regarding how legal

authorities should behave and helps to normatively justify the power that an institution imbues into a power-holder. When authority figures act in normatively appropriate ways, citizens imbue the institution with normativity, which is to say that they lend the institution a moral force that helps to justify its possession of power. Consistent with procedural justice theory, we posit that procedural justice is a positive predictor of normative obligation. If the officer conducting a vehicle stop wields his or her authority in normatively appropriate ways (primarily by being respectful, neutral, accountable, and trustworthy) the citizen could emerge from the encounter with a strengthened sense that the institution is entitled to enforce the law and have their decisions accepted and directives obeyed.

We also assess the role that personal sense of power plays in the context of police-citizen encounters and relations. This construct has received little attention in the legitimacy literature[8] but it is particularly apposite in the current context given the prima facie nature of the two forms of obligation. For the sake of parsimony, we place personal sense of power as a mediator of the treatment, alongside procedural justice, rather than flowing out of procedural justice (like Mentovich 2012, does). We predict that personal sense of power will mediate the treatment effect on normative obligation, and as with procedural justice, the expectation is that induced positive variation in sense of power will be associated with higher average levels of normative obligation. With normative obligation, this would be indicative of active, agentic consent that is rooted in a sense of civic and legal duty.

The pathways from the treatment to willingness to comply with traffic laws that flow through normative obligation reflect the idea that police behaviour can enhance or weaken voluntary deference to those who enforce the law, which in turn can help to motivate voluntarily deference to traffic laws. The pathways from the treatment to willingness to cooperate with the police that flow through normative obligation reflect the idea that police behaviour can enhance or weaken voluntary

[8] Procedural justice may decrease people's sense of power distance regarding the police, because procedural justice has an empowering and/or power equalising quality (Mentovich, 2012). If normative obligation reflects a sense of active and willing consent, one would expect personal sense of power to positively predict normative obligation, with some of the effect of the manipulation on normative obligation going through personal sense of power. If non-normative obligation reflects, among other things, a sense of pragmatism in the face of powerlessness, one would predict that personal sense of power would negatively predict non-normative obligation, and that some of the effect of the effect of the manipulation on non-normative obligation would go through personal sense of power.

deference to those who enforce the law (through procedural justice or personal sense of power). In turn, deference motivates a willingness to cooperate because people are aware that the police want citizens to report crimes and suspicious activity and provide information important to investigation.

Non-normative obligation

By contrast, the dynamics exhibited by non-normative obligation should say something about the effects of less positive police-citizen encounters, as well as whether there is a kind of zero-sum game going on with normative and non-normative forms of obligation. Does, for example, procedural justice policing increase normative obligation and decrease non-normative obligation? If so, how do cooperation and compliance seem to respond to these two shifting motivations?

While theoretical expectations are more tentative for the second form of obligation (as denoted by the dashed lines in Figure 1), we posit a negative relationship between people's experience of procedural justice and non-normative obligation. Legitimacy transforms power into authority, and a positive encounter can help to turn a non-normative form of authority relations into a normative form of authority relations, in which concerns about the consequences of non-obedience are shifted to more active and willing deference to police orders and decisions. By contrast, a negative encounter may increase prudential and/or instrumental obligation. People are unlikely to completely ignore the asymmetrical power relations they have with police officers, so in the relative absence of truly free consent, they would nevertheless be prudent in the face officer demands.

In the context of personal sense of power, we posit that an unfair encounter is associated with lower levels of subjective power, which plausibly then increases one's sense of dull compulsion and fear of the consequences of non-obedience (in other words, prudential obligation to obey). The pathways from the treatment to willingness to comply with traffic laws that flow through decreased procedural justice or sense of power and increased non-normative obligation are expected to diminish the likelihood of legal compliance with the power-holder. Non-normative obligation is also expected to end up with a negative effect on cooperation, because people who feel powerless or afraid of the police are unlikely to come forward voluntarily.

The methodological contribution

Causal mediation analysis is a challenge even with a single mediator, but the nature of the current inquiry makes the issue more complex. We hypothesised a chain of causal pathways where the causal effect is created by the (i) treatment (experimental condition) which is transmitted by (ii) either procedural justice (first model) or sense of power over the police (second model) through (iii) normative/non-normative obligation to obey on (iv) compliance and cooperation.

However, the critical consideration is to isolate what exactly seemed to change as a result of the manipulation and to estimate the relevant causal pathways leading from the treatment to the downstream constructs. By identifying causally mediating effects, we can explain why and how the treatment's effect (a kind of contact which was judged by and large as more negative or more positive) has an effect on the outcome. For example, the effect of treatment (T) on normative obligation is posited to partly run through procedural justice as one mediator (M_1) or personal sense of power as another mediator (M_2). By focusing on the indirect effects, we can answer the question how the treatment influenced the outcome, i.e., to what extent the treatment's effect is attributable to other constructs such as procedural justice or personal sense of power.

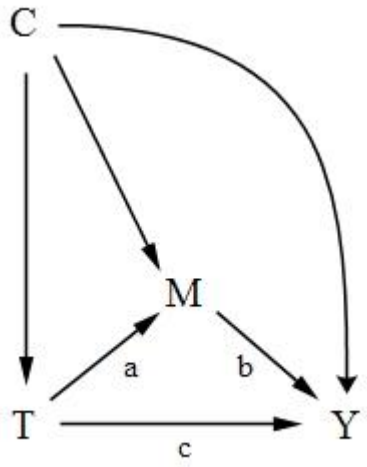


Figure 2 Mediation analysis with a single mediator

Limitations of the traditional approach to causal mediation analysis

Baron and Kenny's (1986) mediation formula is the traditional approach in the social sciences to estimate indirect or mediated effects. Figure 2 depicts an example of a simple case of mediation in an experimental design with a randomised treatment T that has a direct effect on the outcome Y, denoted by c. An indirect effect through the mediator M that is computed as the product of a and b. To control for other variables that were not affected by the treatment, a vector of pre-treatment confounders C.

While this approach is widely used, it has a number of limitations (Mackinnon, Kisbu-sakarya, and Gottschall 2013; Preacher 2015). First, the product method is only applicable when linearity is assumed, which makes the estimation of indirect effects for non-linear models unattainable (Jo 2008). Second, the additivity or no-interaction assumption means that the usual decomposition breaks down in the presence of an interaction between T and M that affects the outcome. In such cases, there is no straightforward way to calculate the total effect (Coffman and Zhong 2012).[9] Finally, scholars sometimes downplay (or completely disregard) the potential pitfalls of causal interpretations regarding the mediated effects. The random assignment of the treatment only guarantees the causal interpretation of the direct effect (T→Y) and the treatment's impact on the mediator (T→M). Critically, it does not shield against the unmeasured confounders of the mediator-outcome relationship (M→Y). Furthermore, there are no easy ways to implement design-based approaches to make the mediator (as-if) randomised (Imai, Tingley, and Yamamoto 2013; Pirlott and Mackinnon 2016).

This paper focusses on some of the potential remedies offered by recent statistical developments on causally ordered mediation. As with all causal techniques, causal mediation analysis requires several identification assumptions that, if satisfied, lend the derived statistical estimates a causal interpretation (Manski 2007). This set of assumptions is usually referred to as sequential ignorability, strong or conditional ignorability (Imai et al. 2011; Pearl 2001). It requires that:

i. The effect of T on Y is unconfounded controlling for C
ii. The effect of M on Y is unconfounded controlling for C and T
iii. The effect of T on M is unconfounded controlling for C

[9] An alternative way of decomposition was offered by Judd and Kenny (1981), nevertheless the currently discussed method goes beyond relaxing the no-interaction assumption offering a more versatile tool than their original proposition.

iv. None of the confounders of M are affected by T

Importantly, randomisation of the treatment only accomplishes (i) and (iii). Assumption (ii) prescribes that there cannot be any unmeasured confounder which would affect the relationship between the mediator and the outcome. This can be achieved either by applying special design-based strategies or relying on an adequate set of pre-treatment covariates C (which were unaffected by the treatment). Finally (iv) demands that there cannot be any confounder of M which was affected by T. Crucially, such a post-treatment confounder would act as a second mediator, providing an alternative conduit transmitting the treatment's effect both through itself and through M. This means that the current set of identification criteria only applies to models with a single mediator, when satisfied allowing researchers to derive generalised natural direct and indirect effects where these effects do have causal properties.

Causal mediation analysis with causally ordered mediators

We have so far considered the single mediator case. Yet, the current application (Figure 1) posits four pathways with two causally ordered mediators, where procedural justice or sense of power are the first intermediate variables, followed by one of the obligation to obey constructs. Figure 3 represents the updated model, where L is the first mediator or post-treatment confounder, and M is the second mediator now transmitting the effects of both T and L. Importantly, the following four pathways emerge with two sequential (or ordered) mediators:

(1) Treatment → Cooperation with police/compliance with the law,
(2) Treatment → Procedural Justice/Sense of power → Cooperation with police/compliance with the law,
(3) Treatment → Normative/non-normative obligation to obey the police → Cooperation with police/compliance with the law,
and finally,
(4) Treatment → Procedural justice/Sense of power → Normative/non-normative obligation to obey the police → Cooperation with police/compliance with the law.

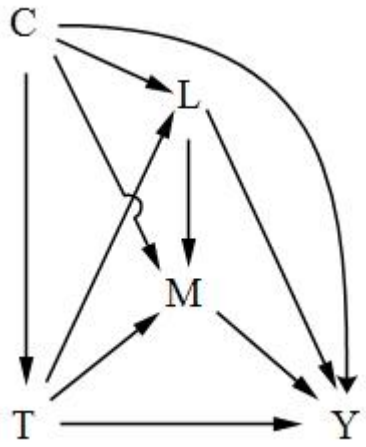


Figure 3 Mediation analysis with two mediators

Yet, according to the sequential ignorability assumption, there can be only a single mediator affected by the treatment (iv). If there are multiple mediators, this assumption is not violated provided that these mediators are causally independent (orthogonal) of each other. This is assumed for procedural justice and sense of power, and normative and non-normative obligation respectively. However, the model used here also posits conditionality. It tests the mediated impact of procedural justice and sense of power on cooperation and compliance through normative and non-normative obligation to obey the police (4). Hence, in the presence of post-treatment confounders, mediated effects are not identifiable based on the previously outlined ignorability criteria.

An elegant resolution of this violation of a key assumption is to shift the focus from single mediators to a vector of mediators (Lange, Rasmussen, and Thygesen 2014; Steen, Loeys, Moerkerke, and Steen 2017; VanderWeele and Vansteelandt 2014). The difference between a mediator M and a post-treatment confounder L is only substantive, otherwise, they are statistically equivalent, which means that any variable affected by treatment T can be added to the vector which will then be robust to unmeasured common causes of various mediators. This approach is not sensitive to the initial ordering of the mediators and even allows for interactions (moderated effects) to be taken into account. Thus, this approach partitions the different pathways to a natural direct effect (NDE) (1) and a (joint) natural indirect effect (NIE) (2)-(4). Estimating the joint effects provides a robust test of the underlying causally mediated mechanism. Another advantage of this approach is that handling the mediators as a

vector only requires a small change in the identification assumptions that need to apply to a vector of mediators instead of one mediator (e.g., procedural justice and normative obligation, or M and L).

One way to estimate these joint effects is to rely on natural effect models and the imputation of the potential outcome (Steen, Loeys, Moerkerke, and Vansteelandt 2017; Vansteelandt, Bekaert, and Lange 2012).[10] When estimating the causal effect, one aims to achieve a comparison of the same participant's chosen values of the outcome variable had that person received the treatment and control condition at the very same moment in time in two hypothetical worlds. In real life, however, we can only observe one of these two outcomes, never both of them. The imputation approach reformulates this problem as an issue of missing data. In the case of a randomised controlled trial, this missingness is addressed by random assignment, which guarantees that the comparison of (marginal) potential outcomes will be unbiased, or exchangeable (i.e., as-if fully observed). This again means that the relationship between the treatment and outcome and the treatment and mediator would not require the use of imputation. However, the same does not apply to the relationship between the mediator and outcome, where this exchangeability can be only assumed conditional on pre-treatment covariates (such as gender, age, etc.) and no uncontrolled confounding. After specifying a model of the outcome variable[11] regressed on the pre-treatment covariates, the treatment, and mediator(s), the values of the potential outcomes will be imputed separately, but simultaneously, because the original dataset does not hold any information regarding their joint distribution, i.e., the unobserved outcome is fully missing (Westreich et al. 2015). In the spirit of this technique and as an added perk, the missing outcome variables are also automatically imputed using the same imputation model. The estimates of NDE and NIE can be obtained upon fitting a (natural effects) mediation model to this imputed dataset (Steen, Loeys, Moerkerke, and Vansteelandt 2017).

Even though these joint effects might shed some light on the grouped overall causally mediated effects of the studied constructs, this approach limits the scope of

[10] Notably, in case of categorical mediators and outcomes the weighting approach yields better results, while for the current, continuous case the imputation-based approach provides more coherent estimation.

[11] One of the most salient advantages of applying natural effect models is that the NDE and NIE can be expressed on a scale that corresponds to the outcome variable, thus providing more straightforward inference.

the analysis, as it does not allow for the assessment of specific pathways ((2), (3), and (4) separately), thus failing to address the initial mediation hypotheses. A further issue is that this perspective does not permit the test of the order of the mediators either. Fortunately, some recent advancements in natural effect models (Steen, Moerkerke, and Vansteelandt, 2017) allow further partitioning of the joint effects. Instead of relying on a two-way decomposition, this approach offers a three-way decomposition in case of two sequential mediators:[12] the NDE remains the same (1), the NIE will now incorporate both (2) and (4), while (3) will be the semi-natural/partial indirect effect (PIE). In other words, the joint natural indirect effect can be partitioned to a natural indirect effect which includes all pathways going through procedural justice/sense of power (L), including the ones that also go through the obligation constructs, and a semi-natural/partial indirect effect, which contains the indirect effect of normative/non-normative obligation (M) that does not go through L. Because there are two Ls and Ms, it needs to be maintained that the pathways through procedural justice and sense of power, and normative and non-normative obligation are none-intertwined, or in other words independent of each other. Furthermore, to identify these effects and in addition to the extended sequential ignorability discussed for the joint effects, two further assumptions need to be met:

v. the effect of L on M is unconfounded controlling for T and C
vi. none of the L and M confounders are affected by T

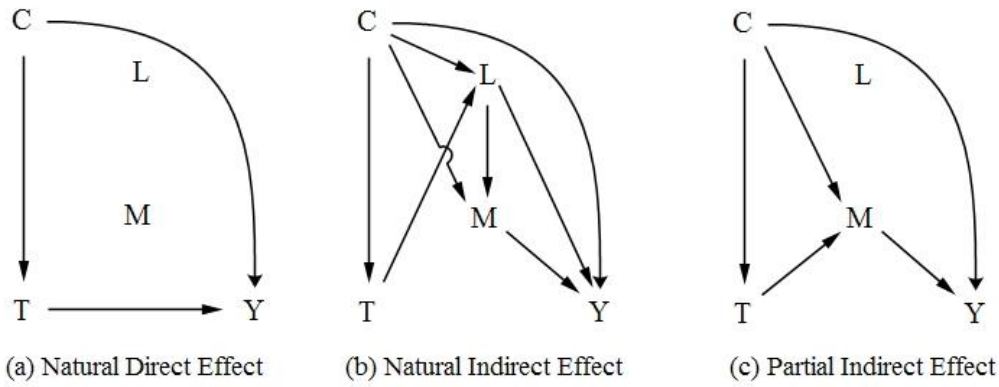


Figure 4 Mediation analysis with two mediators and three-way decomposition

[12] This method can be extended to more than two causally ordered mediators, as shown in Appendix/C.

Figure 4 depicts the decompositions in case of the separate pathways approach. The estimation of the natural indirect and partial indirect effects also requires the imputation of potential outcomes and some further steps that are discussed in Appendix/A.

Method

Design

ScotCET was funded by the Scottish Government to inform the Justice Strategy for Scotland (MacQueen and Bradford 2015). Vehicle stops were conducted during the Scottish Festive Road Safety Campaign 2013/14, with police officers stopping citizens while driving to check the alcohol levels of drivers and conduct routine vehicle safety checks. The 20 road police units involved in the study were divided into 10 matched pairs according to shared geographical characteristics, and within each pair, one unit was randomly assigned to the control group and the other unit was randomly assigned to the treatment group (i.e., block-randomisation). The control group conducted 'business as usual' stops throughout the campaign. The treatment group, following a short 'pre-period' operating 'business as usual', received a combination of verbal and written instruction on how to successfully apply a procedurally just model of policing during routine encounters that aimed to communicate or enable the core aspects of procedural justice: dignity and respect, equality, trustworthy motives, neutrality of decision making, clear explanation, and the opportunity for citizen participation or 'voice'.

ScotCET was designed as a partial replication — and extension — of QCET (Mazerolle et al. 2013). The main objective of QCET was to test the effect of introducing a procedural justice script to police activity in the context of traffic stops in Queensland. In QCET, procedural justice was higher, on average, among the treatment group compared to the control group. Moreover, path analysis indicated that procedural justice explained a good deal of the variation of legitimacy, that the treatment seemed to have an effect on legitimacy, and that procedural justice explained some of that effect (using the standard approach to statistical mediation). There seemed to be a causal effect of the manipulation on legitimacy, some of which seemed to be transmitted via procedural justice.

QCET worked; ScotCET did not. Implementation failure meant that there was a treatment effect but it was in the opposite expected direction (MacQueen and Bradford 2015). A follow-up qualitative study pointed to the possibility that there was a series of communication-based errors in the implementation process that occurred in a context of policing reform and a general perception of organisational injustice within the force (MacQueen and Bradford 2017). These factors seemed to have combined to produce a diffuse, negative effect on the attitudes and behaviours of the officers involved. To give two examples, evidence from the focus groups suggested, first, that officers felt the script impugned their professional integrity and represented an unwarranted intervention into their working lives. This may have affected their performance in an organisational context marked by significant, often unpopular, and certainly poorly-communicated change (the formation of Police Scotland from the eight former regionally-based forced). Put bluntly, officers were already feeling unhappy and being asked to deliver the intervention made them unhappier still. Second, even when the script was implemented properly and by an officer in a positive frame of mind, it may have produced negative outcomes if, for example, it bureaucratised the encounters and/or made them more formal and less natural. Relatedly, there is evidence from the QCET trial that longer encounters can be perceived by citizens as less procedurally fair (Mazerolle et al. 2015); the script may have increased the duration of the interview.

In short, what seems to have happened was not simply a case of implementation failure – which would have most likely lead to 'nil' effects from the experiment – but rather that the design of the intervention, and the way it was communicated to the officers who delivered it, combined to produce a 'knock-on' negative effect on driver perceptions. Recall, however, that Mazerolle et al.'s (2013) focus was not only on the link between treatment and procedural justice, but also on other, downstream constructs like legitimacy, cooperation, and compliance. It is this secondary aspect of ScotCET that we focus on in the current paper, i.e., the impact of the treatment on constructs further down the potential causal chain. This can be done on the current dataset due to the reassessment of the implementation failure, which found no selection-bias and treatment effect consistency and homogeneity (Pósch 2018, Paper 1). In other words, even though we do not possess additional information about what exactly transpired during the police contacts, we know that the differences caused by the treatment are attributable to the research design. Causal mediation analysis permits

to explain how and why the treatment affected the outcomes by relying on the well-defined mediators and focussing on the indirect effects.

Survey and measures

All drivers stopped were issued with a self-completion questionnaire with a prepaid envelope (an online alternative was also offered). Excluding the baseline 'pre-period' of the trial (305 questionnaires returned), 510 questionnaires were returned. In terms of descriptive statistics, broad equivalence between treatment and control groups was achieved. Overall, 63 per cent of respondents were male, the mean age was 50.7 (SD=14.8, min=17, max=87), three quarters (77%) of respondents were homeowners. 41 per cent had a university degree or higher (12 per cent reported holding no qualifications) and the majority were employed (71 per cent), and 73 per cent were married or in a relationship.

Procedural justice was measured by asking research participants whether the officer seemed approachable and friendly, helpful, respectful, professional, fair, and clear in explaining why they had been stopped, whether they trusted the intentions of the officer involved, whether they were confident that the officer was doing the right thing, whether the officer gave them the opportunity to express their views, and whether the officer listened to what they had to say. Response alternatives ranged from 1 'no, not at all' to 4 'yes, completely.'

Sense of power was measured by a single item, 'How much power do you think people like you have over the police?'. Response alternatives ranged from 1 'very little power' to 4 'a lot of power'.

Normative obligation was measured by asking research participants the extent to which they either agreed or disagreed to the following statements: 'I feel a moral obligation to obey the police, ' I feel a moral duty to support the decisions of police officers, even if I disagree with them' and 'I feel a moral duty to obey the instructions of police officers, even when I don't understand the reasons behind them'. Response alternatives were 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree' and 'strongly agree.'

Non-normative obligation was measured by asking respondents the extent to which they agreed or disagreed to the following statements: 'People like me have no choice but to obey the police', 'If you don't do what the police tell you they will treat you badly' and 'I only obey police because I am afraid of them. Response alternatives

were ‘strongly disagree’, ‘disagree’, ‘neither agree nor disagree’, ‘agree’ and ‘strongly agree.’

Willingness to cooperate with the police was captured by asking research participants ‘If the situation arose, how likely would you be to ’: ‘call the police to report a crime you had witnessed’, ‘help police to find someone suspected of a crime by providing information’ and ‘report dangerous or suspicious activities to the police. The response alternatives ranged from 1 ‘not likely at all’ to 4 ‘very likely.’

Willingness to comply with the law was measured ‘All things considered, how likely are you in the future to break the speed limit while out driving’ and ‘All things considered, how likely are you in the future to jump a red light if you are in a hurry.’ The response alternatives ranged from 1 ‘not likely at all’ to 4 ‘very likely.

Results

Scaling

Results from two fitted CFA models using MPlus 7.2 are shown in Table 1 (indicators were set as categorical and all latent constructs were allowed to covary). Included were multiple indicators of procedural justice, normative obligation, non-normative obligation, legal compliance, and willingness to cooperate. Each model also had a single indicator of personal sense of power in its original form, set to be correlated with the latent variables. In the five-factor model, the indicator of non-normative obligation ‘People like me have no choice but to obey the police’ was allowed to cross-load, specifically onto both normative and non-normative obligation (where it loaded negatively on normative obligation and positively on non-normative obligation). This was motivated by Tankebe’s (2013) use of the indicator as a reverse-coded measure of obligation. The exact and approximate fit statistics suggest that the five-factor (M1) fits the data adequately, at least according to the approximate fit statistics, where one typically looks for CFI >.95; TLI >.95; RMSEA <.08 (see Kaplan 2008). The four-factor model, combining normative and non-normative obligation, has a relatively poor approximate fit, at least when judged on the basis of RMSEA and TLI. The Chi-squared statistics are significant in both cases, indicating a bad fit, which is expected when there is a relatively large sample size.

Confirmatory factor analysis models	*Chi-Square*	*df*	*p*	*RMSEA*	*RMSEA 90% CI*	*CFI*	*TLI*
Five-factor model	268	104	<.005	.044	.038; .051	.992	.989
Four-factor model (combining normative obligation and non-normative obligation)	1801	110	<.005	.137	.132; .143	.915	.894

Table 1 Fit statistics for two fitted CFA models

Correlational results

For most of the constructs, factor scores were derived using confirmatory factor analysis (from separate fitted models) and entered into further analysis (except for the single item measure of personal sense of power, which was kept intact). The first column in Table 2 shows that the treatment has a significant weak negative correlation with procedural justice ($r=-0.103$, $p<0.05$), sense of power ($r=-0.113$, $p<0.05$), normative obligation ($r=-0.144$, $p<0.05$), cooperation ($r=-0.094$, $p<0.05$), and compliance ($r=-0.085$, $p<0.1$), and a non-significant negative relationship with non-normative obligation ($r=-0.010$, $p>0.1$). This indicates that being assigned to the treatment condition was negatively related to people's experience of procedural justice in the encounter, their general sense of power during police encounters, their normative obligation to obey, their willingness to cooperate with the police, and their compliance with traffic laws.

As the only construct that was made up of items with negative views of the police, non-normative obligation has a strong negative relationship with procedural justice ($r=-0.463$, $p<0.01$), a moderately strong negative relationship with sense of power ($r=-0.248$, $p<0.01$) and cooperation ($r=-0.279$, $p<0.01$), and a weak negative relationship with normative obligation ($r=-0.061$, $p<0.01$) and compliance ($r=-0.061$, $p<0.01$). This implies that prudential obligation does not seem to be the opposite of consensual obligation. Indeed, the two constructs appear to have a tenuous relationship. Non-normative obligation has a stronger association with procedural justice and sense of power, and with cooperation from the outcome variables.

The rest of the mediating variables follow the expected pattern. Procedural justice has a moderately to strongly positive correlation with the other variables (r=0.223-0.547, p<0.01), as does sense of power (r=0.216-0.547, p<0.01), and normative obligation to obey (r=0.216-0.463, p<0.01). Of the two outcome variables, cooperation has relatively strong relationships with these mediators (r=0.365-0.543, p<0.01), while compliance shows weaker associations with them (r=0.216-0.236, p<0.01). The two outcome variables are weakly associated (r=0.172, p<0.01).

Variables	*Treatment*	*Procedural justice*	*Sense of power*	*Normative obligation*	*Non-normative obligation*	*Cooperation*
Procedural justice	-0.103*					
Sense of power	-0.113*	0.547**				
Normative obligation	-0.144**	0.463**	0.387**			
Non-normative obligation	-0.010	-0.411**	-0.248**	-0.061†		
Cooperation	-0.094*	0.543**	0.410**	0.365**	-0.279**	
Compliance	-0.085†	0.223**	0.216**	0.236**	-0.061†	0.172**

*†p<0.1, *p<0.05, **p<0.01*

Table 2 Correlational results

Natural effect model

In each model, the treatment, mediators, and outcome variables are regressed onto gender, age, housing status, employment, and whether a breath test was conducted during the police encounter. These pre-treatment covariates are not included in either Table 3 or 4 for visual ease, but their coefficients can be found in Appendix/B[13]. The negative values in the tables might be counterintuitive, but they simply indicate the unexpected – and unintended – direction of the treatment effect, which shows that receiving the treatment had detrimental effects on cooperation and compliance. The

[13] As they are pre-treatment variables they take on the same values in each model for the respective outcome.

standard errors shown in Table 3 and 4 are bootstrapped with 1000 replications. Although generally speaking the sum of the partial and natural indirect effects should approximately coincide with the corresponding joint effect of the mediators, some slight discrepancies can arise when modelling with continuous mediators (see details in the Appendix/A).

Willingness to cooperate with the police – Results and Discussion

The first four columns in Table 3 present procedural justice's, sense of power's, normative, and non-normative obligation's separate mediated effect on cooperation with the police. Procedural justice, sense of power, and normative obligation has a weak negative natural indirect effect on different levels of significance (NIE_{coop_pj}=-0.076, $p<0.1$, NIE_{coop_pow}=-0.062, $p<0.1$, NIE_{coop_ob}=-0.079, $p<0.05$) with significant negative natural direct effects of the treatment (NDE_{coop_pj}=-0.168, $p<0.1$, NDE_{coop_pj}=-0.184, $p<0.05$, NDE_{coop_ob}=-0.165, $p<0.1$). The effects of procedural justice and consensual obligation to obey are largely identical, implying that when fitted separately the perceived fair treatment by the police and the felt normative duty to obey the police have a very similar mediated impact on people's willingness to cooperate. Sense of power has a weaker indirect effect, indicating that feeling empowered during a police-citizen encounter increases the likelihood of future cooperation with the police. In contrast, non-normative obligation to obey has a non-significant positive indirect effect with the effect size very close to zero (NIE_{coop_pob}=0.012, $p>0.1$) with a moderately strong significant direct effect (NDE_{coop_pob}=-0.255, $p<0.01$). Thus, prudential obligation does not seem to transmit the treatment's effect on cooperation.

Cooperation	1	2	3	4	5	6	7	8
Procedural justice (NIE)	-0.076† [0.042]							
Sense of power (NIE)		-0.062† [0.032]						
Normative obligation (NIE)			-0.079* [0.038]					
Non-normative obligation (NIE)				0.012 [0.017]				
Pj and ob joint indirect effect					-0.104* [0.051]			
Pj – Natural indirect effect					-0.028 [0.043]			
Ob – Partial indirect effect					-0.073* [0.036]			
Pj and pob joint indirect effect						-0.072 [0.048]		
Pj – Natural indirect effect						-0.066† [0.040]		
POb – Partial indirect effect						-0.010 [0.039]		

Pow and ob joint indirect effect							-0.091*	
							[0.041]	
Pow – Natural indirect effect							-0.041	
							[0.088]	
Ob – Partial indirect effect							-0.047†	
							[0.038]	
Pow and pob joint indirect effect								-0.033
								[0.031]
Pow – Natural indirect effect								-0.043†
								[0.040]
POb – Partial indirect effect								0.011
								[0.091]
Treatment (NDE)	-0.168†	-0.184*	-0.165†	-0.255**	-0.140†	-0.171*	-0.155†	-0.214*
	[0.089]	[0.091]	[0.089]	[0.096]	[0.087]	[0.085]	[0.086]	[0.091]

*†p<0.1, *p<0.05, **p<0.01, the squared brackets straddle the bootstrapped standard errors*

pj=procedural justice, pow=sense of power, ob=free/normative obligation to obey the police, pob=prudential/non-normative obligation to obey the police

Table 3 Natural effect models with two causally ordered mediators for cooperation with the police

The fifth column shows the joint, natural indirect and partial indirect effects of procedural justice and consensual obligation to obey the police. Their joint effect is significant, going slightly below -0.1 (JIE_{coop_pjob}=-0.104, p<0.05), with an also significant direct effect (NDE_{coop_pjob}=-0.140, p<0.1). When the causal ordering is considered, normative obligation to obey is capable of reserving most of its mediated effect (PIE_{coop_ob}=-0.73, p<0.05), solely transmitting the impact of the treatment, as procedural justice does not have a significant natural indirect effect (NIE_{coop_pjob}=-0.028, p>0.1). This provides further support to the original hypothesis, indicating that normative obligation on its own is capable of mediating the previous contact's (treatment) impact on willingness to cooperate, even without the procedural justice's and normative obligation's jointly mediated effect.

The joint effect of procedural justice and non-normative obligation to obey (shown in column six) is not significant (JIE_{coop_pjpob}=-0.072, p>0.1), with a significant direct effect of the treatment (NDE_{coop_pjpob}=-0.171, p<0.05). Prudential obligation to obey the police has a non-significant partial indirect effect on cooperation (PIE_{coop_pob}=-0.010, p>0.1) which takes the opposite direction compared to its sole NIE. Procedural justice's significant natural indirect effect (NIE_{coop_pjpob}=-0.066, p<0.1) alludes that when the police are perceived as fair and neutral it increases the willingness to cooperate with the police.

The seventh column, shows that sense of power and normative obligation have a significant joint effect on willingness to cooperate (JIE_{coop_powob}=-0.091, p<0.05). Further decomposition indicates that consensual obligation to obey's significant partial indirect effect (PIE_{coop_powob}=-0.047, p<0.05) mediates the effect of the treatment, while sense of power has a non-significant natural indirect effect with a comparable effect size (NIE_{coop_powob}=-0.041, p>0.1). This implies that while the pathway that only goes through legitimacy does mediate the effect of the treatment, the pathways going through personal sense of power do not. The previous contact's direct effect remains significant here as well (NDE_{coop_powob}=-0.155, p<0.1).

Finally, column eight shows the non-significant joint effect of sense of power and non-normative obligation to obey the police (JIE_{coop_powpob}=-0.033, p>0.1). Non-normative obligation to obey's natural indirect effect is virtually unchanged compared to its sole NIE, with a weak non-significant positive effect (PIE_{coop_powpob}=0.011, p>0.1), while sense of power has a smaller, but significant negative natural indirect effect (NIE_{coop_powpob}=-0.043, p<0.1). These findings seem to mirror the ones found for procedural justice and non-normative obligation, insofar as prudential obligation does

not appear to transmit the treatment's effect on willingness to cooperate with the police, while the pathways going through the first mediator do. The effect of the treatment remains significant here as well (NDE_{coop_powpob}=-0.214, $p<0.05$).

An overview of the results suggests that the joint effects of duty to obey and either of the two first mediators (procedural justice or sense of power) have the strongest indirect effect on cooperation. Further decomposition shows that – as theorised earlier – consensual obligation to obey almost fully transmits the effects of the treatment on cooperation, while the pathways going through either of the first two mediators are non-significant in the same models (model 5, 7). Moreover, after the three-way decomposition only normative obligation to obey's mediated effect remained significant on the 5% level in one of the models. In contrast, non-normative obligation to obey didn't appear to be significant in any of the models. It follows that adding non-normative obligation to obey at best has no impact, at worst, marginally diminishes the mediated impact of procedural justice and sense of power, producing non-significant joint, but significant natural indirect effects. All in all, coercive obligation does not transmit the impact of contact with the police, while in the same models the pathways going through procedural justice and sense of power do. Thus, the statistical evidence seems to support that from the obligation constructs only normative obligation has a causally mediated effect on people's willingness to cooperate with the police.

Conspicuously, after considering the causal ordering, procedural justice and sense of power only remained significant in the models with prudential obligation (model 6, 8). Yet, one would expect these variables having a natural indirect effect on willingness to cooperate in all models, at least due to the joint pathways also going through consensual obligation to obey the police. Unfortunately, the decomposition pursued by the current paper does not allow to determine whether the pathway going through both duty to obey and one of the first mediators would provide significant results.

Furthermore, it is important to acknowledge the persistently significant natural direct effects' effect sizes which consistently surpassed the strength of the indirect effects in all models. This indicates that the selected variables do not fully mediate the relationship between the treatment and cooperation with the police, or in other words, that certain portion of the variation in the experiences with the police remains unaccounted for by the mediators, and that the candidate mediators are only imperfect conduits of the effect of the treatment on willingness to cooperate.

Compliance	1	2	3	4	5	6	7	8
Procedural justice (NIE)	-0.061† [0.039]							
Sense of power (NIE)		-0.069† [0.040]						
Normative obligation to obey (NIE)			-0.086* [0.041]					
Non-normative obligation (NIE)				0.001 [0.008]				
Pj and ob joint indirect effect					-0.113* [0.055]			
Pj – Natural indirect effect					-0.055† [0.036]			
Ob – Partial indirect effect					-0.058* [0.017]			
Pj and pob joint indirect effect						-0.078† [0.048]		
Pj – Natural indirect effect						-0.058† [0.035]		
POb – Partial indirect effect						-0.016 [0.037]		

Pow and ob joint indirect effect							-0.099†	
							[0.052]	
Pow – Natural indirect effect							-0.039†	
							[0.022]	
Ob – Partial indirect effect							-0.060*	
							[0.018]	
Pow and pob joint indirect effect								-0.058
								[0.040]
Pow – Natural indirect effect								-0.056†
								[0.040]
POb – Partial indirect effect								-0.012
								[0.042]
Treatment (NDE)	-0.096	-0.088	-0.071	-0.158	-0.043	-0.079	-0.076	-0.118
	[0.112]	[0.123]	[0.119]	[0.126]	[0.113]	[0.117]	[0.117]	[0.125]

*†p<0.1, *p<0.05, **p<0.01, the squared brackets straddle the bootstrapped standard errors*

pj=procedural justice, pow=sense of power, ob=free/normative obligation to obey the police, pob=prudential/non-normative obligation to obey the police

Table 4 Natural effect models with two causally ordered mediators for compliance with the law

Compliance with the law – Results and discussion

The first four columns of Table 4 show a very similar picture to Table 2. Procedural justice's, sense of power's, and normative obligation to obey's NIEs are significant (NIE_{compl_pj}=-0.061, p<0.1, NIE_{compl_pow}=-0.069, p<0.1, NIE_{compl_ob}=-0.086, p<0.05) with non-significant NDEs (NDE_{compl_pj}=-0.096, p>0.1, NDE_{compl_pow}=-0.088, p>0.1, NDE_{compl_ob}=-0.071, p>0.01). All the effect sizes are relatively weak, procedural justice's and sense of power's NIEs having a very similar magnitude, with an edge to normative obligation to obey, which indicates that consent plays a little more important role in influencing compliance than the perceived fair treatment and decision making by the police or the feeling of being empowered. Yet again, non-normative obligation to obey has an almost non-existent mediated effect (NIE_{compl_pob}=0.001, p>0.1) with a non-significant direct effect (NDE_{compl_pob}=-0.158, p>0.1). Therefore, prudential obligation does not seem to transmit the treatment's effect towards compliance.

Column five displays the joint and causally ordered effects of procedural justice and normative obligation to obey. The joint effect is moderately strong (JIE_{compl_pjob}=-0.133, p<0.05) with a non-significant direct effect (NDE_{compl_pjob}=-0.043, p>0.1). After taking into account the causal sequence, both the partial indirect effect of consensual obligation (PIE_{compl_ob}=-0.058, p<0.05) and the natural indirect effect of procedural justice (NIE_{compl_pj}=-0.055, p<0.1) remain significant. The effect sizes of both variables are reduced compared to their sole NIE, and unlike with cooperation, normative obligation alone only partially mediates the treatment's impact on compliance. This implies that both the normative obligation to obey (partial indirect effect) and the pathways through perception of police fairness (natural indirect effect) influence people's compliance with the law. Normative obligation's imperfect transmission suggests that when it comes to compliance people do not simply rely on their recognition of the power of the authorities and their faith of the rightfulness of their jurisdiction (legitimacy), but they also consider their expectations regarding the fairness and due process promised by the police (procedural justice).

The joint effect of procedural justice and non-normative obligation (column six) is significant (JIE_{compl_pjpob}=-0.078, p<0.1) with a non-significant direct effect (NDE_{compl_pjpob}=-0.079, p>0.1). Examining the causal ordering implied that procedural justice's natural indirect effect is significant (NIE_{compl_pjpob}=-0.058, p<0.1), while prudential obligation's partial indirect effect is not (PIE_{compl_pob}=-0.016, p>0.1). This exemplifies the usefulness of the three-way decomposition, which here shows that

procedural justice's natural indirect effect is the only significantly mediating effect for both cooperation and compliance, which manages to turn the joint effect significant in case of the latter, but not the former. By contrast, non-normative obligation to obey is once again unable to transmit the treatment's effect, entailing that compliance with the law seems to be unrelated to being intimidated by the police.

In the seventh column, the joint effect of sense of power and normative obligation to obey emerges as significant (JIE_{compl_powob}=-0.099, $p<0.1$) with significant natural indirect and partial indirect effects for sense of power and free obligation to obey respectively (NIE_{compl_pjob}=-0.039, $p<0.1$, PIE_{compl_pjob}=-0.060, $p<0.05$). In accordance with the model for cooperation, normative obligation to obey appears to only partially mediate the treatment's effect on compliance, this time with sense of power being also significant. This indicates that both the willing authorisation and the feeling of having power over the police mediate the previous police contact's impact on compliance with traffic laws. Thus, both the psychological understanding of being recognised during the encounter and the motivational force of police legitimacy contribute to compliance with the established laws. The natural direct effect of the treatment is not significant here either (NDE_{compl_pjob}=-0.043, $p>0.1$).

Lastly, the joint effect of sense of power and non-normative obligation is also examined (column eight), which is non-significant (JIE_{compl_powpob}=-0.058, $p>0.1$) with a non-significant direct effect (NDE_{compl_powpob}=-0.118, $p>0.1$). Crucially, when considering the sequence of the mediators, only the natural indirect effect of sense of power is significant (NIE_{compl_powpob}=-0.056, $p<0.1$), while prudential obligation only transmits a very small portion of the effect of the treatment amounting to a non-significant relationship (PIE_{compl_powpob}=-0.012, $p>0.1$). This result also closely resembles the one perceived for cooperation: the non-normative form of obligation to obey does not seem to be influential in predicting the desired outcome, while the pathways through sense of power only have a weak effect.

From the models fitted for compliance with the law, the joint effect of procedural justice and normative obligation to obey stands out as the strongest, followed by the joint effect of sense of power and normative obligation to obey, and the sole NIE of normative obligation to obey. Markedly, and in contrast with the models for cooperation, the sequential approach reveals that normative obligation does not solely mediate the impact of the treatment and that the pathways through both procedural justice and sense of power retain a significant impact on compliance.

Hence, willingness to comply with traffic laws is not only influenced by normative obligation, but also by people's understanding of the fairness of the police, and their felt grasp of power in police encounters. In line with the earlier findings for cooperation, non-normative obligation does not mediate the effects of earlier contact with the police. This raises further doubt whether policing that strengthens instrumental forms of obligation can influence either cooperation with the police or compliance with the law in either direction. Finally, the natural direct effects of the treatment are non-significant which implies that the mediators managed to successfully transmit a substantial share of the influence of contact with the police.

Conclusion

The substantive goal in this paper was to consider the possibility—raised by Tankebe and colleagues (Bottoms and Tankebe 2012; Tankebe 2013; cf. Johnson et al. 2014)—that citizens could interpret standard measures of legitimacy in ways that go beyond what is intended. The status quo is that when surveys have asked people whether they "should obey the police even when they think the police are wrong," agreement to survey measures is typically taken to signify the view that the police are legitimate and that they have the moral right to expect obedience from citizens. Yet, Tankebe and others have argued that positive answers that survey respondents give to these questions could also partly reflect the view that it is dangerous to defy the police, or that one has little choice but to be obedient and comply with officer instructions. While prior studies have found that duty to obey is highly correlated with (a) the belief that police officers are procedurally fair, (b) the judgement that the institution is appropriate, proper and just (Jackson et al. 2012; Sunshine and Tyler 2003; Tyler 2006b; Tyler and Fagan 2008; Tyler and Huo 2002) and (c) the willingness to proactively cooperate with the police, the danger remains that standard survey methods conflate two different forms of obligation.

This is important because, given the centrality of internalised duty to obey to the notion of legitimacy, researchers need to be confident that the measurement tools are only capturing truly free consent and not, in addition, pragmatic or strategic compliance from people who believe they lack the power to resist. For instance, Sun et al.'s (2017) study set in a coastal Chinese city assumed that standard measures of obligation to obey conflate normative and non-normative motives, and as a result placed obligation outside of the legitimacy construct. The substantive aim of our study

was to address the open and empirical question of whether it is possible to capture normative obligation to obey. We have added to the available conceptual discussion and methodological resources by assessing how two new scales of obligation operate in a procedural justice framework. The first was designed to tap into normative obligation (the items stress a moral duty to obey the commands of officers) while the second was designed to tap into non-normative obligation (the items stress compliance through fear of reprisal and/or dull compulsion). Applied to a real-world setting of road policing in Scotland, the design of the RCT allowed us to test the effect of altering the dynamics of police-citizen encounters on the outcomes of legal compliance and willingness to cooperate with effects mediated through (a) immediate outcomes (i.e., experienced fairness and sense of power over the police) and (b) distal outcomes (i.e., normative and non-normative obligation to obey). Studying these downstream causal effects of the treatment effect is essential to inform both the procedural justice theory and police practice.

What, then, have we learnt from this study? In the context of road policing in Scotland, at least, normative and non-normative obligation seem to be distinct and largely unrelated constructs. Normative obligation seems to be a sense of active consent rooted in the experience of fair treatment, fair decision-making, the provision of voice, and the belief that the officer had trustworthy intentions, while non-normative obligation seems to be a resistant sense of dull compulsion to the restriction of freedom that police officers can represent. Exploiting the fact that the RCT induced variation in the experience of procedural fairness and personal sense of power (albeit in small *and expectantly negative ways*), we found that normative obligation operated very differently to non-normative obligation when included in a procedural justice model. First, normative obligation was positively (and strongly) correlated with key theoretical variables, including experienced procedural justice, personal sense of power, willingness to cooperate, and legal compliance. Second, non-normative obligation was negatively correlated with procedural justice, sense of power and cooperation. Third, and despite their similar correlates, the normative and non-normative aspects of obligation to obey showed only a very weak, tenuous association, implying that the two constructs are largely independent of each other. Fourth, normative obligation to obey emerged as the most important causal mediator for willingness to cooperate, and one of the important mediators for compliance with the law. Fifth, and by contrast, non-normative obligation did not mediate the treatment's

impact on either of the outcomes, implying that non-normative considerations do not channel the previous contact's impact on legal compliance and cooperation with the police. Finally, procedural justice and sense of power only mediated the treatment's effect on cooperation in the absence of normative obligation in the model, whilst they always mediated the impact on legal compliance in addition to normative obligation.

One has, of course, to define *a priori* legitimacy as the right to power and the entitlement to be obeyed (Tyler 2006a, 2006b; Tyler and Jackson 2013) but if one elaborates the concept in this way, then the *entitlement to be obeyed* aspect of legitimacy does seem to accord with the normative obligation captured in the present RCT given its grounding in fair and legitimate authority relations. So long as duty to obey is defined and measured in a way that stresses truly free consent, one could reasonably include obligation in the legitimacy concept. In other words, if one defines legitimacy as the right to power and the authority to govern (in the eyes of citizens), then including a sense of moral duty to obey in the operational definition makes sense. We, therefore, recommend, on the basis of our findings, that scholars use measures of normative obligation that stress the phrase *moral duty*. Legitimacy is about power relations; authorities make claims about their right to dictate appropriate behaviour and to have their directives obeyed (Bottoms and Tankebe 2012), and the reception of these claims amongst citizens plausibly involve some *direct* sense of authorisation and deference (if citizens accept these claims) and a rejection of their right to expect compliance (if citizens do not accept these claims). Yet, researchers do need to be confident that the survey questions tap solely into a sense of obligation that centres upon rights and responsibilities in the context of legal authority. We thus recommend the use of measures that stress the notion of 'moral duty.'

Lastly, causal mediation analysis appeared to be an effective tool to distinguish between the treatment's direct and indirect effects. Natural effects models, in particular, are flexible as they can be applied even in case of non-linear modelling and in the presence of treatment-mediator interactions that might affect the outcome. We recommend that in future studies where several sequential mediators are present one should carry out the analytical steps outlined earlier. First, researchers should examine the mediated effect with a single mediator, then the joint effects, and only after that turn to further decomposition. Decomposing the effects to natural and partial indirect effects is crucial because often times this is the only way to identify which mediator is actually influential in the model. Ultimately, causal mediation analysis models always

need to be informed by the existing literature and judged against the causal identification criteria.

Limitations of the analysis

Although the three-way decomposition presented here can help to unpack the underlying causal mechanisms, some difficulties still prevail when interpreting the various effects. The partial indirect effects are straightforward, as they represent the pathway going through only the second mediator towards the outcome. By contrast, the natural indirect effects incorporate both the pathway going through the first and first and second mediators (i.e., their jointly mediated effect). This means that the interpretation of the natural indirect effects for procedural justice and sense of power is murkier than it would be desirable, and future studies might want to seek finer decomposition to elucidate the effects (Daniel et al. 2015).

Possibly the most serious limitation of the current study is the strong assumption of causal independence made for the pairs of normative and non-normative obligation to obey, and procedural justice and sense of power respectively. This assumption is fundamentally untestable, yet it determines the viability of the presented effect decomposition, and the causal claims made throughout the article. For normative and non-normative obligation we believe that there are strong reasons to assume that these constructs are independent of each other. The correlational evidence shows that they have a very weak barely significant relationship. Moreover, they appear to be functionally different: while consensual obligation seems to channel the effect of the treatment to the outcome variables, prudential obligation does not seem to transmit the same effect. Yet, other results make us more cautious regarding this claim, as both obligation items have fairly similar bivariate relationships with procedural justice and sense of power. Further studies are needed to establish whether this assumption of independence can be justified.

As far as procedural justice and sense of power are concerned, our assumption stands on an even shakier ground. Correlational results imply that these constructs are strongly related to each other. Moreover, there are other competing theoretical models from the one presented here which might be equally plausible. It is possible for instance that procedural justice informs how people evaluate their personal sense of power, which in turn influences their ideas about the legitimacy of the police. Such a model would require three causally ordered mediators (i.e., procedural justice → sense

of power → normative/non-normative obligation to obey). Thus, to check the robustness of our results, we also ran natural effects models to test this proposition, which produced very similar results to the ones presented here (see: Appendix/C), suggesting that loosening this assumption would not substantially change the conclusions drawn here.

Thoughts on future research

Our findings support the idea that, when police officers act in ways that accord with normative expectations regarding fair inter-personal treatment and decision-making, this can help to create a sense among those they interact with, the police are generally fair and that the institution is legitimate and entitled to be obeyed (see also Cheng 2015; White, Mulvey, and Dario 2016). Scotland is a country with relatively low crime rates and little history of the sort of tense and fraught police-citizen relations that one can find in some other parts of the world. In a country like this, people may tend to interpret the measures of normative obligation in the way that is intended by researchers. This may not be the case in a country like Ghana or Brazil, or indeed in certain communities in large Metropolitan cities in the US. We encourage research in other parts of the world to see if similar findings emerge.

One of the biggest remaining questions is about the dynamics of non-normative obligation to obey. The current findings imply that normative and non-normative considerations might have very different downstream effects (Tyler et al. 2015). These findings chime with studies on the perception of procedural injustice that appear to influence outcomes very differently than procedural justice (Augustyn 2016; Reisig, Mays, and Telep 2018). They also contribute to our understanding of the potentially contrasting nature of normative and instrumental authority-relations (Anderson, John, and Keltner 2012; Mentovich 2012).

Finally, the natural effect models used here are only one approach in the big family of methods of causal mediation analysis. There are semi-parametric alternatives which allow for post-treatment confounding (Imai and Yamamoto 2013), g-computation solutions which can be used for sequentially ordered mediators and post-treatment confounding alike (Daniel et al. 2015; De Stavola et al. 2015), and so on. We hope that the current example will encourage other researchers who want to estimate causal indirect effects to immerse themselves in similar methods to the one applied here.

Appendix/A – Natural effect models with two causally ordered mediators – technical appendix

As discussed in detail by Steen, Moerkerke, and Vansteelandt (2017), the natural effects (partial indirect effect, natural indirect effect, and natural direct effect) for two causally ordered mediators can be derived following six analytical steps.

1. First, two models need to be fitted for the two mediators. In these models the first mediator L is conditional on the treatment (T) and pre-treatment covariates (C), while the second mediator M is conditional on T, C, and L.
2. Second, a model has to be fitted for the outcome (Y) which is conditional on T, C, L, and M.
3. Third, the analysed data set has to be replicated four times with three auxiliary variables created for the values of the treatment (denoted as t, t', and t''). For the first replication, t will contain the observed values of T=t for each individual, and for the second replication t will take on the counterfactual values of 1-T, whilst t' and t'' will include the observed treatment levels for both of these replications. The third and fourth replication will be a duplicate of these two extended data sets which will however include the counterfactual values for t' and t''.
4. Fourth, the ratio-of-mediator probabilities (i.e., densities) of either of the mediators is computed for each row in the data set, which will be used as weights in subsequent analysis (for details see: Lange et al., 2014). However, these weights are prone to being unstable in case of continuous variables, which might lead to less precise natural effect estimates and possible finite sample bias (Steen et al., 2016). As a consequence, sometimes the sum of the natural and partial indirect effects' effect sizes can slightly differ from the joint effect's effect size estimates. As either of the mediators' weights can be used, this article always relied on the first mediator's (procedural justice's or sense of power's) weights.
5. Fifth, the model outlined in the second step is used for the imputation of the nested counterfactuals as fitted values for every row in the extended data set.
6. Finally, the natural effect model under scrutiny is fitted to the extended data set where the imputed outcomes are regressed on the values of t, t', and t'', and C, while weighting by one of the weights derived at step four.

Appendix/B – Table of covariate effects for the two different models

It follows from the estimation method described in the article, and detailed in the technical appendix, that the coefficients of the covariates' will be the same regardless of the number of mediators included for the two models fitted for the two outcome variables respectively. As shown by Table 1a, being female (β_{female_coop}=0.209, p<0.05) and higher levels of educational attainment (β_{educ_coop}=0.083, p<0.1) were associated with the readiness to cooperate with the police, with none of the other covariates showing a significant positive relationship. For compliance, being female (β_{female_compl}=0.274, p<0.05) and being retired instead of being unemployed/in non-traditional employment ($\beta_{retired\_compl}$=0.763, p<0.01) were significantly positively associated with people's reported compliance with traffic laws, while being married showed a negative association ($\beta_{married\_compl}$=0.763, p<0.05). None of the other covariates had a significant relationship with compliance with law.

	Cooperation	*Compliance*
Female (vs male)	0.209*	0.274*
	[0.085]	[0.122]
Age (years)	0.006	0.005
	[0.004]	[0.006]
Breath test (vs no breath test)	-0.089	-0.173
	[0.099]	[0.143]
Married (vs not married)	0.013	-0.334**
	[0.088]	[0.129]
Educational attainment	0.083†	0.029
	[0.049]	[0.072]
Employed (vs other)	0.177	0.133
	[0.191]	[0.279]
Retired (vs other)	0.236	0.763*
	[0.246]	[0.335]
House owner (vs other)	-0.059	-0.423
	[0.234]	[0.310]
Renter (vs other)	0.014	-0.062
	[0.249]	[0.322]

*†p<0.1, *p<0.05, **p<0.01*

Table 1/a Pre-treatment covariates in the respective natural effect models for cooperation with the police and compliance with the law

As a test of treatment effect heterogeneity we defined interactions with the pre-treatment covariates and the treatment, but as none of those effects reached statistical significance or showed substantial effect sizes we did not include them in our final models.

Appendix/C – Natural effect models with three causally ordered mediators – decomposition, results

As noted in the article's discussion, there are plausible alternative theoretical models which might warrant further consideration. As a test of one of them, it is posited that previous experience with the police has an impact on the beliefs regarding the procedural justice of the police, which in turn influences the public's sense of power during police encounters, which then affects the considerations regarding the legitimacy of the police, finally impacting the willingness to cooperate with the police and compliance with the law (see Figure 1a). Without going into detail regarding the modified identifying assumptions and altered method of estimation (details on them can be found in Steen, Moerkerke, and Vansteelandt (2017)), the number of pathways towards the outcome increases from the previous four to eight:

1. Treatment → Cooperation/Compliance
2. Treatment → Procedural justice → Cooperation/Compliance
3. Treatment → Sense of power → Cooperation/Compliance
4. Treatment → Normative/Non-normative obligation → Cooperation/Compliance
5. Treatment → Procedural justice → Sense of power -> Cooperation/Compliance
6. Treatment → Procedural justice → Normative/Non-normative obligation → Cooperation/Compliance
7. Treatment → Sense of power → Normative/Non-normative obligation → Cooperation/Compliance
8. Treatment → Procedural justice → Sense of power → Normative/Non-normative obligation → Cooperation/Compliance

Following the earlier estimation strategy, here a four-way decomposition is feasible, where (1) will provide the natural direct effect, (2, 5, 6, 8) the natural indirect effect of procedural justice, (3, 7) the partial natural indirect effect of sense of power, and (4) the partial indirect effect of normative/non-normative obligation to obey. Noticeably, the natural indirect effect will encompass all pathways that go through the first mediator, the partial natural indirect effect all paths that go through the second,

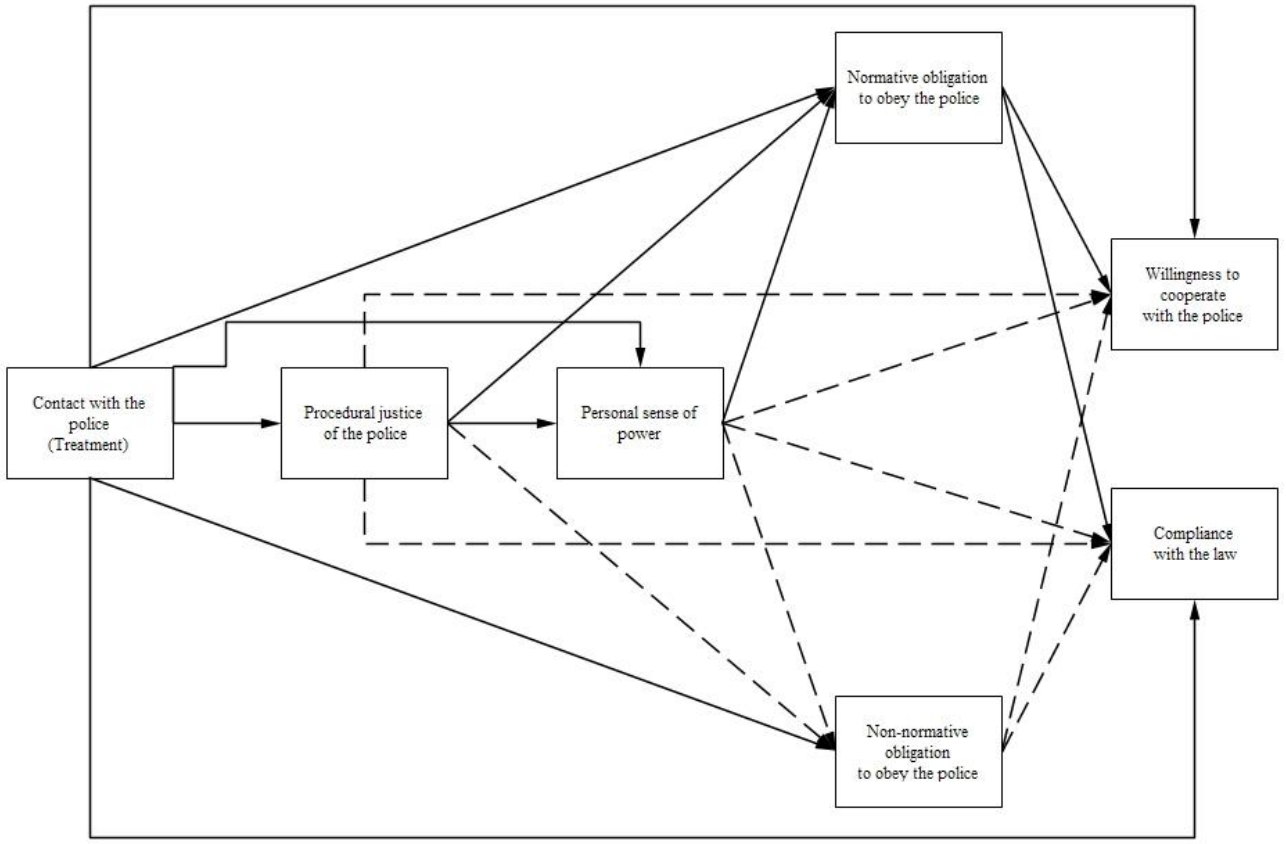


Figure 1a An alternative theoretical model of cooperation and compliance with three sequentially ordered mediators

but not the first mediator, and the partial indirect effect only the one where the third mediator is present on its own.

The results for three causally ordered mediators is presented in Table 2a. The first column contains the results for normative obligation to obey as the third mediator and willingness to cooperate as the outcome, and finds a significant joint indirect effect ($JIE_{coop\_pjpowob}$=-0.108, p<0.05) with a non-significant direct effect ($NDE_{coop\_pjpowob}$=-0.138, p>0.1). From the three indirect effects only the partial indirect effect of normative obligation to obey is significant ($PIE_{coop\_pjpowob}$=-0.079, p<0.05), but neither the partial natural indirect effect of sense of power ($PNIE_{coop\_pjpowob}$=-0.019, p>0.1), nor the natural indirect effect of procedural justice ($NIE_{coop\_pjpowob}$=-0.014, p>0.1) are significant. These results mirror the ones found in Table 3 where only consensual obligation to obey had a mediated impact on cooperation.

	Cooperation		*Compliance*	
Pj, pow, and ob joint effect	-0.108*		-0.107†	
	[0.051]		[0.057]	
Pj natural indirect effect	-0.014		-0.029	
	[0.049]		[0.046]	
Pow partial natural indirect effect	-0.019		-0.032	
	[0.020]		[0.036]	
Ob partial indirect effect	-0.079*		-0.054*	
	[0.034]		[0.013]	
Pj, pow, and pob joint effect		-0.164†		-0.095†
		[0.088]		[0.056]
Pj natural indirect effect		-0.096*		-0.064†
		[0.046]		[0.034]
Pow partial natural indirect effect		-0.067†		-0.032†
		[0.035]		[0.020]
Pob partial indirect effect		-0.005		-0.009
		[0.036]		[0.040]
	-0.138	-0.082†	-0.069	-0.081
	[0.085]	[0.047]	[0.118]	[0.116]

*†p<0.1, *p<0.05, **p<0.01, the squared brackets straddle the bootstrapped standard errors*
pj=procedural justice, pow=sense of power, ob=free/normative obligation to obey the police, pob=prudential/non-normative obligation to obey the police

Table 2a Natural effects models with three causally ordered mediators for cooperation with the police and compliance with the law

The second column has non-normative obligation as the third mediator with cooperation as the outcome variable. It reports a significant joint ($JIE_{coop\_pjpowpob}$=-0.164, p<0.1) and direct effects ($NDE_{coop\_pjpowpob}$=-0.082, p<0.1). Notably, this joint effect has the strongest effect size among all the models for cooperation. From the mediators the pathways going through procedural justice ($NIE_{coop\_pjpowpob}$=-0.096, p<0.05) and sense of power ($PNIE_{coop\_pjpowpob}$=-0.067, p<0.1) emerge as significant, while non-normative obligation has a non-significant partial indirect effect ($PIE_{coop\_pjpowpob}$=-0.005, p>0.1). Yet again, these results are very similar to the ones found in the earlier table for cooperation (Table 3).

In the third column the third mediator is normative obligation, with legal compliance as the outcome variable. The joint effect is significant ($JIE_{compl\_pjpowob}$=-0.107, p<0.1), with a non-significant direct effect ($NDE_{compl\_pjpowob}$=-0.069, p>0.1). From the rest of the effects only the partial indirect effect of consensual obligation to obey is significant ($PIE_{compl\_pjpowob}$=-0.054, p<0.05), while neither the PNIE for sense of power ($PNIE_{compl\_pjpowob}$=-0.032, p>0.1), nor the NIE for procedural justice ($NIE_{compl\_pjpowob}$=-0.029, p>0.1) are. These results are slightly different compared to the ones found in Table 4 where the natural indirect effects were also significant.

Finally, the fourth column's third mediator is non-normative obligation with legal compliance as the outcome variable. This model has a significant joint effect ($JIE_{compl\_pjpowpob}$=-0.095, p<0.1) and a non-significant direct effect ($NDE_{compl\_pjpowpob}$=-0.081, p>0.1). Both the partial natural indirect effect of sense of power ($PNIE_{compl\_pjpowpob}$=-0.032, p<0.1) and natural indirect effect of procedural justice ($NIE_{compl\_pjpowpob}$=-0.064, p<0.1) are significant, with a non-significant partial indirect effect for prudential obligation ($PIE_{compl\_pjpowpob}$=-0.009, p>0.1). These effects also follow a very similar pattern to the one found in Table 4.

References

Anderson, Cameron, Oliver P. John, and Dacher Keltner. 2012. "The Personal Sense of Power." *Journal of Personality* 80(2):313–44.

Anon. 2015. *The President's Task Force on 21st Century Policing*.

Augustyn, Megan Bears. 2016. "Updating Perceptions of (In)Justice." *Journal of Research in Crime and Delinquency* 53(2):255–86.

Baron, Reuben M. and David a. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6):1173–82.

Baron, Reuben M. and David A. Kenny. 1986. "Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6):173–82.

Bottoms, Anthony and Justice Tankebe. 2012. "Beyond Procedural Justice: A Dialogic Approach To Legitimacy in Criminal Justice." *Journal of Criminal Law & Criminology* 102(1):119–70.

Bradford, Ben, Kristina Murphy, and Jonathan Jackson. 2014. "Officers as Mirrors." *British Journal of Criminology* 54(4):527–50.

Coffman, D. L. and W. Zhong. 2012. "Assessing Mediation Using Marginal Structural Models in the Presence of Confounding and Moderation." *Psychological Methods* 17(4):642–64.

Daniel, R. M., B. L. De Stavola, S. N. Cousens, and S. Vansteelandt. 2015. "Causal Mediation Analysis with Multiple Mediators." *Biometrics* 71(1):1–14.

Hamm, J. A., R. Trinkner, and J. D. Carr. 2017. "Fair Process, Trust, and Cooperation: Moving Toward an Integrated Framework of Police Legitimacy." *Criminal Justice and Behavior* 44(9):1183–1212.

Huq, A., T. Tyler, and S. Schulhofer. 2011. "Mechanisms for Eliciting Cooperation in Counter Terrorism Policing: Evidence from the United Kingdom." *Journal of Empirical Legal Studies* 8(4):728–61.

Huq, Aziz Z., Tom R. Tyler, and Stephen J. Schulhofer. 2011. "Why Does the Public Cooperate with Law Enforcement? The Influence of the Purposes and Targets of Policing." *Psychology, Public Policy, and Law* 17:419–50.

Imai, K., D. Tingley, and T. Yamamoto. 2013. "Experimental Designs for Identifying

Causal Mechanisms." *Journal of the Royal Statistical Society Series A-Statistics in Society* 176(1):5–51.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105(4):765–89.

Imai, Kosuke and Teppei Yamamoto. 2013. "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments." *Political Analysis* 21(2):141–71.

Jackson, Jonathan et al. 2012. "Why Do People Comply with the Law?" *British Journal of Criminology* 52(6):1051–71.

Jackson, Jonathan. 2018. "Norms, Normativity, and the Legitimacy of Justice Institutions: International Perspectives." *Annual Reveiw of Law and Social Sciences* 14 In pres.

Jackson, Jonathan and Jacinta M. Gau. 2015. "Carving up Concepts? - Differentiating between Trust and Legitimacy in Public Attitudes towards Legal Authority." Pp. 49–69 in *Interdiisciplinary Perspectives on Trust - Towards Theoretical and Methodological Integration*, edited by E. Shockley, T. M. S. Neal, L. PytlikZillig, and B. Bornstein. Springer.

Jo, Booil. 2008. "Causal Inference in Randomized Experiments With Mediational Processes." *Psychological Methods* 13(4):314–36.

Johnson, Devon, Edward R. Maguire, and Joseph B. Kuhns. 2014. "Public Perceptions of the Legitimacy of the Law and Legal Authorities: Evidence from the Caribbean." 48(4):947–78.

Jonathan-Zamir, Tal and Amikam Harpaz. 2018. "Predicting Support for Procedurally Just Treatment: The Case of the Israel National Police." *Criminal Justice and Behavior* 45(6):840–62.

Judd, Charles M. and David A. Kenny. 1981. *Estimating the Effects of Social Interventions*. Camdridge University Press.

Kaplan, David. 2008. *Structural Equation Modeling - Foundations and Extensions*. 2nd ed. SAGE.

Lange, Theis, Mette Rasmussen, and Lau Caspar Thygesen. 2014. "Assessing Natural Direct and Indirect Effects through Multiple Pathways." *American Journal of Epidemiology* 179(4):513–18.

Mackinnon, David P., Yasemin Kisbu-sakarya, and Amanda C. Gottschall. 2013. "Developments in Mediation Analysis Oxford Handbooks Online Developments in Mediation Analysis." Pp. 1–28 in *Oxford Handbook of Quantitative Methods*, vol. 2, edited by T. D. Little. New York: Oxford University Press.

MacQueen, Sarah and Ben Bradford. 2015. "Enhancing Public Trust and Police Legitimacy during Road Traffic Encounters: Results from a Randomised Controlled Trial in Scotland." *Journal of Experimental Criminology* 11(3):419–43.

MacQueen, Sarah and Ben Bradford. 2017. "Where Did It All Go Wrong? Implementation Failure—and More—in a Field Experiment of Procedural Justice Policing." *Journal of Experimental Criminology* 13(3):321–45.

Manski, Charles F. 2007. *Identification for Prediction and Decision*. Harvard University Press.

Mazerolle, Lorraine et al. 2015. "Optimising the Length of Random Breath Tests: Results from the Queensland Community Engagement Trial." *Australian & New Zealand Journal of Criminology* 48:256–76.

Mazerolle, Lorraine, Emma Antrobus, Sarah Bennett, and Tom R. Tyler. 2013. "Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice." *Criminology* 51(1):33–63.

Mentovich, Avital. 2012. *The Power of Fair Procedures - The Effect of Procedural Justice on Perceptions of Power and Hierarchy*. New York University.

Murphy, K., B. Bradford, and J. Jackson. 2016. "Motivating Compliance Behavior Among Offenders: Procedural Justice or Deterrence?" *Criminal Justice and Behavior* 43(1):102–18.

Murphy, Kristina and Adrian Cherney. 2012. "Understanding Cooperation with Police in a Diverse Society." *British Journal of Criminology* 52(1):181–201.

Murphy, Kristina, Tom R. Tyler, and Amy Curtis. 2009. "Nurturing Regulatory Compliance: Is Procedural Justice Effective When People Question the Legitimacy of the Law?" *Regulation and Governance* 3(1):1–26.

Pearl, Judea. 2001. "Direct and Indirect Effects." Pp. 411–20 in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence UAI'01*.

Pirlott, Angela G. and David P. Mackinnon. 2016. "Design Approaches to Experimental Mediation ☆." *Journal of Experimental Social Psychology* 66:29–38.

Pósch, Krisztián. 2018. *Prying Open the Black Box of Causality: A Causal Mediation Analysis Test of Procedural Justice Policing*. Retrieved (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3087872).

Preacher, Kristopher J. 2015. "Advances in Mediation Analysis: A Survey and Synthesis of New Developments." *Annual Review of Psychology* 66:825–52.

Reisig, M. D. and C. Lloyd. 2008. "Procedural Justice, Police Legitimacy, and Helping the Police Fight Crime: Results From a Survey of Jamaican Adolescents." *Police Quarterly* 12(1):42–62.

Reisig, Michael D., Ryan D. Mays, and Cody W. Telep. 2018. "The Effects of Procedural Injustice during Police–citizen Encounters: A Factorial Vignette Study." *Journal of Experimental Criminology* 14(1):49–58.

Slocum, Lee Ann, Stephanie Ann Wiley, and Finn-Aage Esbensen. 2016. "The Importance of Being Satisfied." *Criminal Justice and Behavior* 43(1):7–26.

De Stavola, Bianca L., Rhian M. Daniel, George B. Ploubidis, and Nadia Micali. 2015. "Mediation Analysis with Intermediate Confounding: Structural Equation Modeling Viewed through the Causal Inference Lens." *American Journal of Epidemiology* 181(1):64–80.

Steen, Johan, Tom Loeys, Beatrijs Moerkerke, and Johan Steen. 2017. "Flexible Mediation Analysis with Multiple Mediators." *American Journal of Epidemiology* 186(2):184–93.

Steen, Johan, Tom Loeys, Beatrijs Moerkerke, and Stijn Vansteelandt. 2017. "Medflex : An R Package for Flexible Mediation Analysis Using Natural Effect Models." *Journal of Statistical Software* 76(11):1–45.

Sun, Ivan Y., Yuning Wu, Rong Hu, and Ashley K. Farmer. 2017. "Procedural Justice, Legitimacy, and Public Cooperation with Police: Does Western Wisdom Hold in China?" *Journal of Research in Crime and Delinquency* 54(4):454–78.

Sunshine, Jason and Tom R. Tyler. 2003. "The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing." *Law and Society Review* 37(3):513–48.

Tankebe, Justice. 2009. "Public Cooperation with the Police in Ghana: Does Procedural Fairness Matter?" *Criminology* 47(4):1265–93.

Tankebe, Justice. 2013. "Viewing Things Differently: The Dimensions of Public Perceptions of Police Legitimacy." *Criminology* 51(1):103–35.

Tyler, Phillip Atiba Goff, and Robert J. MacCoun. 2015. "The Impact of Psychological

Science on Policing in the United States: Procedural Justice, Legitimacy, and Effective Law Enforcement." *Psychological Science in the Public Interest* 16(3):75–109.

Tyler, T., J. Fagan, and A. Geller. 2014. "Street Stops Police Legitimacy: Teachable Moments in Young Urban Men's Legal Socialization." *Journal of Empirical Legal Studies* 11(14):751–85.

Tyler, T. R. and Y. J. Huo. 2002. *Trust in the Law - Encouraging Public Cooperation With the Police and the Courts*. New York: Russell Sage Foundation.

Tyler, Tom and Jeffrey Fagan. 2008. "Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities?" *Ohio State Journal of Criminal Law* 6:231–75.

Tyler, Tom R. 2003. "Procedural Justice, Legitimacy, and the Effective Rule of Law." *Crime and Justice* 30:283–357.

Tyler, Tom R. 2004. "Enhancing Police Legitimacy." *The Annals of the American Academy of Political and Social Science* 593:84–99.

Tyler, Tom R. 2006a. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57:375–400.

Tyler, Tom R. 2006b. *Why People Obey the Law*. Princeton: Princeton University Press.

Tyler, Tom R. and Jonathan Jackson. 2013. "Future Challenges in the Study of Legitimacy and Criminal Justice." Pp. 83–104 in *Legitimacy and Criminal Justice - An International Exploration*, edited by J. Tankebe and A. Liebling. Wiley.

Tyler, Tom R. and Jonathan Jackson. 2014. "Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation, and Engagement." *Psychology, Public Policy, and Law* 20(1):78–95.

Tyler, Tom R., Jonathan Jackson, and Avital Mentovich. 2015. "The Consequences of Being an Object of Suspicion - Potential Pitfalls of Proactive Police Contact." *Journal of Empirical Legal Studies* 12(4):602–36.

Tyler, Tom R., Stephen J. Schulhofer, and Aziz Z. Huq. 2010. "Legitimacy and Deterrence Effects in Counterterrorism Policing: A Study of Muslim Americans." *Law and Society Review* 44:365–402.

VanderWeele, Tyler J. and Stijn Vansteelandt. 2014. "Mediation Analysis with Multiple Mediators." *Epidemiologic Methods* 2(1):95–115.

Vansteelandt, Stijn, Maarten Bekaert, and Theis Lange. 2012. “Imputation Strategies for the Estimation of Natural Direct and Indirect Effects.” *Epidemiologic Methods* 1(1):131–58.

Westreich, Daniel et al. 2015. “Imputation Approaches for Potential Outcomes in Causal Inference.” *International Journal of Epidemiology* 44(5):1731–37.

White, Michael D., Philip Mulvey, and Lisa M. Dario. 2016. “Arrestees’ Perceptions of the Police.” *Criminal Justice and Behavior* 43(3):343–64.

Wolfe, Scott E., Justin Nix, Robert Kaminski, and Jeff Rojek. 2016. “Is the Effect of Procedural Justice on Police Legitimacy Invariant? Testing the Generality of Procedural Justice and Competing Antecedents of Legitimacy.” *Journal of Quantitative Criminology* 32(2):253–82.

Interlude 3

Paper 3 provides the fullest test of the comprehensive model outlined in the introduction. Using natural effects models, we estimated the causal pathways from previous contact with the police to two societally desirable outcomes, compliance with the law and cooperation with the police. In these sequentially mediated models, normative obligation to obey emerged as the primary mediator, making the other two preceding mediators (procedural justice and sense of power) often non-significant. The robustness of these findings is augmented by the nature of the modelling where the effects jointly mediated by multiple mediators are assigned to the first mediator in the sequential order. In other words, normative obligation to obey remained the most important mediator even after its jointly mediated effect was incorporated into the indirect effect of procedural justice or sense of power. Notably, this free duty to obey was the only mediator transmitting the impact of contact on cooperation with the police, although contact retained a significant direct effect. Moreover, normative obligation to obey had the strongest indirect effect when mediating the impact of the treatment on compliance with the law. This implies that for both cooperation with the police and compliance with the law, consideration of consent to police actions was the most important causal mechanism when transmitting the impact of the previous contact with police officers.

By contrast, non-normative obligation did not mediate the causal influence of contact on either of the outcomes. In the presence of non-normative obligation, the other two mediators (procedural justice and sense of power) were significant, transmitting the effect of contact on the outcomes. The correlation with free duty to obey that converged towards zero calls into question whether this prudential aspect of obligation is really the other side of the same coin or something that is outside of the remit of the procedural justice framework.

If indeed police legitimacy is the most important conduit of the influence of contact on cooperation and compliance, a more detailed analysis comparing the two aspects of legitimacy, free duty to obey and normative alignment, might provide a better insight regarding this mediating role. In Paper 4, two experiments manipulate procedural justice and respect for boundaries to assess to what extent these two police legitimacy constructs mediate their impact on willingness to cooperate with the police.

Paper 4: Testing Complex Social Theories with Causal Mediation Analysis and G-Computation: Towards a Better Way to Do Causal Structural Equation Modelling

Krisztián Pósch

Abstract

Complex social scientific theories are conventionally tested using linear structural equation modelling (SEM). However, the underlying assumptions of linear SEM often prove unrealistic, making the decomposition of direct and indirect effects problematic. Recent advancements in causal mediation analysis can help to address these shortcomings, allowing for causal inference when a new set of identifying assumptions are satisfied. This paper reviews how these ideas can be generalised to multiple mediators, with a focus on the post-treatment confounding and causal ordering cases. Using the potential outcome framework as a rigorous tool for causal inference, the application is the theory of procedural justice policing. Analysis of data from two randomised experiments shows that making similar parametric assumptions to SEMs and using g-computation improves the viability of effect decomposition. The paper concludes with a discussion of how causal mediation analysis improves upon SEM and the potential limitation of the methods.

Keywords: causal mediation analysis, causal ordering, structural equation modelling, g-computation, causal inference, police legitimacy, potential outcome framework, post-treatment confounding, procedural justice policing, sensitivity analysis

"Only when they must choose between competing theories do scientists behave like philosophers." (Thomas S. Kuhn)

Introduction

The social sciences are full of relatively complex theories that involve direct and indirect causal pathways. For example, Rivera and Tilcsik's (2016) survey experiment tested whether higher class signals in résumés, mediated by the applicants' perceived fit and commitment to the job, influenced whether the respective male or female participant was invited for an interview, using multi-group linear structural equation modelling (SEM) to examine the indirect (mediated) effects. Many social scientific researchers – especially those in sociology, psychology and criminology – rely on linear SEMs for testing theories of similar complexity, and this technique has several advantages in that, for example, it provides global and (in some cases) comparative model fit estimates and permits simultaneously the fitting of complicated measurement and structural models without aggregation of measurement error (e.g., Kaplan 2008; Tomarken and Waller 2005). Linear SEMs can also be expanded to accommodate data structures of a multilevel and/or longitudinal nature. A further testament to the popularity of linear SEMs is that the two most cited articles in *Sociological Methods & Research* (Bentler and Chou 1987; Browne and Cudeck 1992) were also written on this very subject.

For the social sciences to accumulate a robust body of knowledge that provides credible policy prescriptions, researchers need to test the causality of their often times relatively convoluted models. But many researchers seem – on the surface at least – to be unaware of the difficulties within such endeavours. Rivera and Tilcsik's (2016) article is commendable, as it attempted to causally assess their hypotheses, but they did not test the causal identification assumptions of SEM (Bollen and Pearl 2013; Keele 2015b), nor did they draw upon the methodological literature on causal mediation analysis (Keele 2015a; VanderWeele 2015, 2016). As emphasised by Kenny (2008), mediation analysis is a form of causal analysis where disregarding the underlying causal assumptions can lead to misspecified models and thus misleading results.

To address these difficulties, this paper examines the methodological challenges within and potential of causal mediation analysis with multiple mediators. Presenting some causal alternatives to linear SEMs, it draws upon papers by De

Stavola et al. (2015) and Daniel et al. (2015), both of which use SEM with g-computation to consider statistical issues of post-treatment confounding and causal ordering in causal mediation analysis. Here, these two techniques are discussed with a focus on the interpretation of the results and necessary identification assumptions (for technical details regarding the estimation and modelling, please see the cited papers) and, as a motivating example, the theory of procedural justice policing is used (Tyler 2006; Jackson et al. 2012). This paper provides a comprehensive overview of the different approaches available for causal mediation analysis with multiple mediators and goes beyond a recent publication on causal mediation analysis (VanderWeele 2015); both techniques considered here were devised contemporaneously to this book's publication. The aforementioned two methods improve upon the traditional linear SEM approach in at least two ways: (1) they rely on the potential outcome framework as a rigorous tool to make the causal identification assumptions explicit, and devise formal definitions of the direct and indirect effects and (2) these methods allow for more flexible modelling by loosening some of the parametric requirements, thus providing weaker, and more attainable assumptions, than the ones required for linear SEM.

This article is organised as follows. The first section focusses on a central prediction of procedural justice theory: namely, that people's judgements on the legitimacy of the police mediate the effects of their perceptions of police procedural justice and legality on their willingness to cooperate with the police. Two experiments are outlined: the first manipulated the perceived procedural justice of the police, and the second manipulated the perceived legality of the police. The second section discusses how linear SEMs traditionally derive mediated effects and highlights some of the potential pitfalls of this approach. The third section briefly reviews causal mediation analysis with a single mediator. The fourth section discusses the different approaches researchers can take when working with multiple mediators. The fifth and sixth sections discuss two particular instances of complex social scientific theories: mediators with post-treatment confounding and causally ordered mediators, and the findings from the two experiments are presented. The paper concludes with a consideration of the findings, an outline of some of the limitations of the methods, and some recommendations for applied researchers.

Procedural justice policing and the legitimacy of the police

The theory of procedural justice is built on the idea that when people evaluate their interactions with the police, they are primarily focussed on whether or not the officer (a) makes objective and neutral decisions, and (b) treats them in a fair and respectful manner. When the police act in procedurally just ways, citizens feel that their input is considered, their status in the community is affirmed, and that the police as an institution has legitimate authority (Mazerolle et al. 2013; Tyler 2006; Tyler, Goff, and MacCoun 2015; Tyler and Jackson 2014). In addition to procedural justice, the concept of 'bounded authority' has recently been introduced into the police legitimacy framework (Huq, Jackson, and Trinkner 2017; Trinkner, Jackson, and Tyler 2017). Bounded authority captures the idea that people expect authority figures to respect the limits of their rightful authority, for example, that police officers do not act as if they are above the law or do not become involved in situations that they have no right to be in. People divide their lives into domains, and in each of these domains they put a cap on how much interference from the legal authorities they can tolerate, and this boundary condition can shape their judgements on the legitimacy of that authority.

Legitimacy judgements have two constitutive elements: the right to power and the authority to govern (Tyler 2006; Tyler and Jackson 2013). Applied to the police, right to power judgements can be operationalised as institutional trust (the belief that institutional actors can be trusted to wield their power appropriately, where trust constitutes the normative justifiability of power) or as normative alignment (the belief that institutional actors respect key societal norms regarding how they should behave, where normativity constitutes the normative justifiability of power). Authority to govern activates the moral duty to accept the right of the police to make decisions and dictate appropriate behaviour, prompting voluntary consent and obedience because of the source rather than the content (Bradford 2014; Tyler and Jackson 2013). Even though the building blocks of legitimacy (normative alignment and duty to obey) are usually agreed upon, their relationship is sometimes debated: some argue that these two elements mutually reinforce each other (e.g., Hough et al. 2013) while others claim that normative alignment is a predictor of duty to obey (e.g., Huq, Jackson, and Trinkner 2017). A good deal of research (Tyler and Fagan 2008; Tyler et al. 2015; White, Mulvey, and Dario 2016) has shown that legitimacy has an impact on certain socially desirable outcomes, and among these outcomes, willingness to cooperate with the police is the focus of this paper.

To test procedural justice theory, two experiments manipulate people's perceptions of procedural justice or the legality of the police through descriptions of fake police encounters. Study 1 (n=215) and Study 2 (n=235) were conducted in July 2013 in two subsequent weeks on the Amazon Mechanical Turk website with participants from the United States. Both studies used a similar newspaper article about police roadside checks, manipulating either the perceived procedural justice or legality of the police. It is assumed here that procedurally unjust and illegal treatments have a negative impact on how people form their attitudes about police legitimacy, which in turn transmits their impact on willingness to cooperate with the police.

Procedural justice of the police, legality of the police, normative alignment with the police, obligation to obey the police, and willingness to cooperate with the police were measured with three items each on a 1-5 Likert-scale almost exclusively with construct-specific response alternatives. Gender, age, ethnicity, and state of origin were also measured. For further details regarding the experiments and procedure please refer to Appendix/A.

Structural equation modelling and the traditional definition of indirect effects

Throughout this paper SEM refers to the traditional linear models that most researchers use, rather than certain recent developments in the field that have yet to become standard (e.g., Liu et al. 2014; Mayer et al. 2017; Sardeshmukh and Vandenberg 2016). SEM conventionally relies upon the product method (Baron and Kenny 1986) to estimate mediated effects. Let us assume that we have a treatment (T) that channels its effect (partially) through a mediator (M) for the outcome (Y). The direct effect is T's unmediated impact on Y; the mediated effect is the product of the estimates for T's effect on M and M's effect on Y. Therefore, the name of the product method refers to how the point estimate of the mediated effect is derived. Crucially, the presence and absence of direct and indirect effects is determined by the significance of the effects, although effect sizes should still be considered even in case of non-significant coefficients.

Despite the widespread appeal of this approach, the product method has four limitations. First, SEMs posit *effect homogeneity*, that is, that every unit in the population has the same causal effect. This is an untestable and unrealistic assumption on the population level, as it must pertain to each individual (Robins and Greenland 1992). Second, the product method can only identify the total effect as the sum of the

direct and indirect effects in the absence of a treatment-mediator moderated effect that would influence the outcome (*no-interactions*) (Imai et al. 2011; Imai, Keele, and Yamamoto 2010a). The presence of such a treatment-mediator interaction would be a clear sign that effect homogeneity is violated (Kline 2015). However, even if the effect homogeneity assumption can be relaxed, it remains unclear where to assign the interaction effect. This leads to the failure of the decomposition and makes the direct and indirect effects inextricable (Mackinnon 2008; Mackinnon, Kisbu-sakarya, and Gottschall 2013). A third limitation is that the product method requires the *linearity assumption*, which should apply not only to the outcome variable but to all variables including the mediator(s) (Jo 2008). This linearity assumption guarantees effect constancy, that is, that the effect of one variable on another will be independent of the level of a third variable. Conversely, in non-linear systems the chosen level of M would influence the effect of T on Y, thus prohibiting the additivity of effects (Pearl 2014).

The final limitation concerns not the method itself, but its application. Users of SEM often hope to answer causal questions, and one of the key assumptions to guarantee this is that there are no omitted influential variables (i.e., unmeasured confounders). *Yet, even if the treatment T is randomly assigned, only the T-Y and T-M relationships are randomised, while the M-Y relationship is not*. Rivera and Tilcsik's (2016) study is instructive here. They assumed that the randomised treatment's mediated effects can be deemed causal, when in fact, the effects might be influenced by an unmeasured confounder[14] (Judd and Kenny 1981; VanderWeele 2015).

A brief review of causal mediation analysis with a single mediator

Causal mediation analysis with a single mediator helps to address the limitations mentioned earlier by making the causal identifying[15] assumptions more explicit. Moreover, it also helps to overcome them by permitting non-parametric identification

[14] To further complicate the matter, a simple random assignment of the mediator is not feasible, as even in such a case M needs to remain an outcome of T, but due to the randomisation would become unaffected by T (Coffman and Zhong 2012; Imai, Keele, and Yamamoto 2010b; Luke Keele 2015a). Hence special designed based strategies need to be employed to address this problem (Imai, Tingley, and Yamamoto 2013; Pirlott and Mackinnon 2016).

[15] Identification throughout the paper refers to causal identification, while in the SEM literature it usually alludes to model-based identification. The test of these identification criteria always preludes the statistical analysis as a necessary but not sufficient step of causal analysis. Moreover, this identification permits the calculation of the effects of interest irrespective of the chosen statistical model for estimation (Manski 2007; Moerkerke et al. 2015).

and incorporating the treatment-mediator interaction whilst still allowing the decomposition of effects. For the new definitions of direct and indirect effects the potential outcome framework can be used. At the heart of this framework is a thought experiment in which (assuming a binary treatment for the sake of simplicity) a person receives both the treatment and control simultaneously at the same point in time. Naturally, a person can only receive one of these conditions and we can never observe what would have happened had this person been assigned to the other condition. Nevertheless, provided that the preconceived assumptions are satisfied, the two outcomes are estimable on the population level. The potential outcome framework treats certain counterfactual values as missing, and the only way to address this missingness is to rely on identifying assumptions regarding these unobservable quantities (Keele 2015b; Westreich et al. 2015). If these identifying assumptions are met, they will permit the estimation of population level causal effects[16].

For causal mediation analysis, the *sequential ignorability assumption*[17] was proposed (Imai, Keele, and Tingley 2010; Imai, Keele, and Yamamoto 2010; Pearl 2001). This states that for a treatment T, a mediator M, and an outcome Y with T=t and M=m and controlling for a vector of pre-treatment covariates C, there is:

i. No unmeasured confounding of the T-Y relationship or $Y_{tm} \perp\!\!\!\perp T|C$
ii. No unmeasured confounding of the M-Y relationship also given T or $Y_{tm} \perp\!\!\!\perp M|C,T$
iii. No unmeasured confounding of the T-M relationship or $M_t \perp\!\!\!\perp T|C$
iv. No unmeasured M-Y confounder L that was affected by T or $Y_{tm} \perp\!\!\!\perp M_{t*}|C$

As indicated earlier, *random assignment of T only satisfies (i) and (iii)* from the four assumptions. The first three assumptions are conventional 'no unmeasured

[16] Some scholars (e.g., Daniel et al. 2015; Preacher 2015; De Stavola et al. 2015; Wang and Arah 2015) add consistency and the stable unit treatment value assumption (SUTVA) as further requirements. However, as discussed by Shadish, Cook, and Campbell (2002) SUTVA is a more general fundamentally design based assumption. Moreover, Pearl (2010) argues that the consistency assumption is in fact a theorem required by all assumptions stated in the potential outcome framework. Even if neither SUTVA nor consistency are included explicitly, they will be presumed for all causal analysis discussed in the paper.

[17] Notably, Pearl (2014) advocated milder assumptions and argued that sequential ignorability is a sufficient, but not necessary assumption for identifying causal effects. Imai and Keele (2015) contested his propositions and argued for the more stringent requirements discussed by this article.

confounding' assumptions, while the fourth invokes the 'cross-world independence' assumption where t and t* stand for two values of the treatment we wish to compare (e.g., in case of a binary treatment t=1 is the treatment and t*=0 is the control). Crucially, this cross-world independence assumption also prescribes that *there can be only a single mediator affected by the treatment* (no post-treatment confounder L).

The sequential ignorability assumption permits the definition of new effects. Overall, the conditional expectations for a particular outcome will take the form of $E[Y_t,M_{(t^*)}]$ where t and t* are set at a freely chosen level of the treatment for Y and M. The *controlled direct effect (CDE)* only requires assumptions (i) and (ii), and considers a specified value of M=m and captures the expected increase in Y when T changes from T=0 to T=1. This is a direct effect, since the effect of T is not transmitted through M. The value of CDE might change depending on the chosen value of m:

$$\text{CDE}(m)=E[Y(1,m)-Y(0,m)] \quad (1)$$

Both natural effects require all assumptions (i-iv) to be estimable. The *natural direct effect (NDE)* is similar to the controlled direct effect, as it estimates the expected increase in Y when T changes from T=0 to T=1, but it does not hold m constant; instead it permits m to take its value in the 'natural' way for each individual as if that individual had been assigned to the control condition:

$$\text{NDE}=E[Y(1,M(0))-Y(0,M(0))] \quad (2)$$

The *natural indirect effect (NIE)* does the opposite of NDE as it approximates the expected increase in Y when the treatment is kept at T=1, while M is freed to take its natural value of m for the treatment and the control group respectively. This is an indirect effect that captures the effect of T on Y that is transmitted through M:

$$\text{NIE}=E[Y(1,M(1))-Y(1,M(0))] \quad (3)$$

Finally, the *total effect (TE)* can be decomposed to the sum of the NDE and NIE:

$$\text{TE}=E[Y(1)-Y(0)]= \quad (4)$$

$$\{E[Y(1,M(1))-Y(1,M(0))]\}+$$
$$\{E[Y(1,M(0))-Y(0,M(0))]\}=$$
$$NIE+NDE$$

As described above, identification of the direct and indirect effects through the potential outcome framework does not posit the no-interaction assumption, which allows for the effect decomposition even in the presence of such an association. Moreover, it is non-parametrically identified, hence it does not require the effect homogeneity or linearity assumptions, either of which permits more flexible modelling. For an example of how these effects are estimated using g-computation, please refer to Appendix/B.

Causal mediation analysis with multiple mediators

As with SEMs, causal mediation analysis always starts with a qualitative stage of model building. This stage is inherently theoretical – as alluded to in the quote at the beginning of this article – with the researcher distilling knowledge about prior scholarship, the research design, and potential temporal order to logically structure the theoretical model (Bollen and Pearl 2013). As expressed by (iv), if more than one mediator is present, the sequential ignorability assumption might be violated, threatening the identifiability of the NDE and NIE. Thus, in the presence of multiple mediators, there are four different strategies an analyst can consider: *assume causal independence, model the joint effects, assume post-treatment confounding,* or *assume sequential ordering*. The decision regarding the appropriate strategy cannot be data-driven, it has to be informed by the researcher's knowledge regarding the existing literature.

When *mediators are causally independent of one another* (i.e., parallel or non-intertwined) the same analytical strategy can be pursued as with a single mediator. Notably, this causal independence is an untestable assumption that makes it difficult to assess whether Figure 1 (a) or (b) is more suitable for the constructs analysed. Nonetheless, an obvious way of examining the potential dependence between variables is to regress T, C, and L on the M of interest. Significant relationships can provide a reasonable indication that the variables are dependent on each other (Imai and Yamamoto 2013). However, even if no statistically significant association emerges, it is usually difficult to argue for the orthogonality of the mediators, unless there is a

convincing theoretical reason to do so, or such orthogonality is artificially created (e.g., through varimax rotation in exploratory factor analysis). A rare example of such independence was provided by Taguri, Featherstone, and Cheng (2018), who examined two unrelated techniques to prevent dental cavities, through antibacterial and fluoride therapy mediators. Nevertheless, it is usually difficult to encounter such clear-cut cases in the social sciences. Finally, assessing the mediators one at a time will also fail if there are interactions between the effects of the various mediators on the outcome (Lange, Rasmussen, and Thygesen 2014).

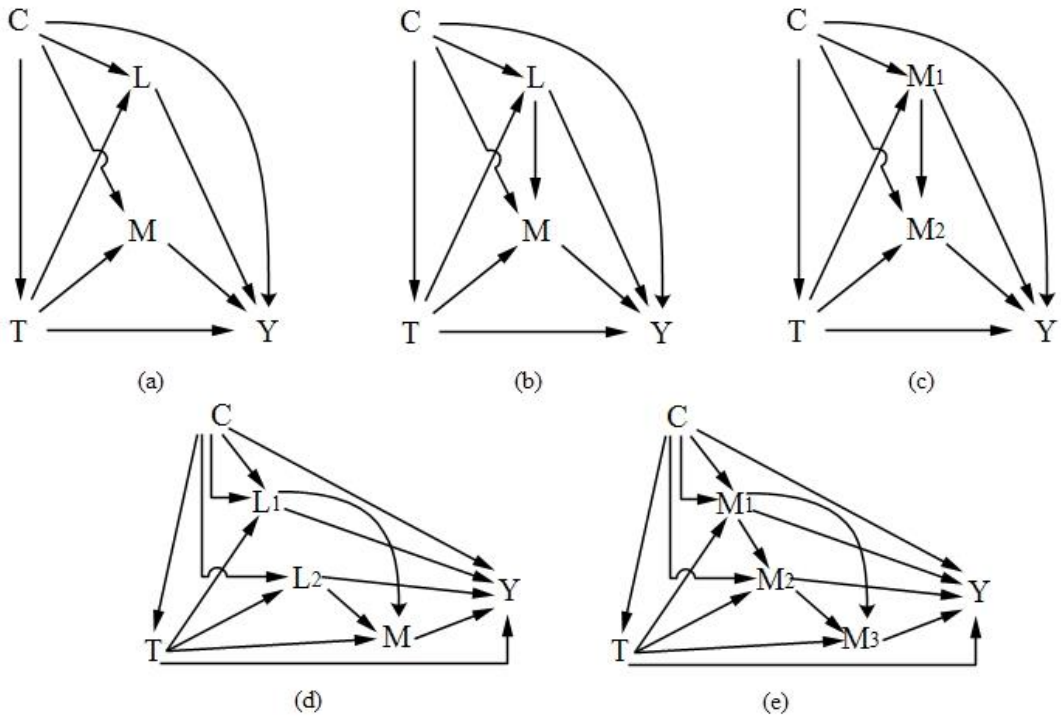


Figure 1 Mediation analysis (a) with two causally independent mediators, (b) with post-treatment confounding where M is dependent on L, (c) with two sequential mediators M_1 and M_2, (d) with post-treatment confounding where M is dependent on L_1 and L_2, (e) with three sequential mediators M_1, M_2 and M_3

Assuming causal dependence between L and M, a simple solution is to *examine their joint effect and treat them as a vector of mediators* (Steen et al. 2017b; VanderWeele and Vansteelandt 2014). Importantly, L and M are statistically equivalent, the post-treatment confounder label of L is only substantive, hence, L can

be considered as a second mediator. Handling multiple mediators as a single vector is robust to unmeasured common causes of various mediators, and can even accommodate cases with causal ordering of M_1 and M_2 similar to Figure 1 (c). Crucially, the sequential ignorability assumption (i-iv) is still valid, but now for a vector of mediators instead of a single mediator. Admittedly, this approach limits the scope of the analysis, yet it can be pragmatic for certain research questions (e.g., if we are only interested in the mediated effects of legitimacy as a whole on the outcome of interest). Even so, for many applied cases this technique will remain untenable. For instance, it would be hard to justify this approach in Rivera and Tilcsik's (2016) study where the commitment and fit for a job are fairly different aspects. Similarly, this strategy would not allow one to test the unique impact of moral alignment with the police and duty to obey the police. In such circumstances, assuming causal mediation analysis with post-treatment confounders or sequentially ordered mediators need to be considered, which will be discussed in the next two sections.

Causal mediation analysis with post-treatment confounding

Avin, Shpitser, and Pearl (2005) in their proof showed that conditioning on L does not permit the non-parametric identification of natural effects, but only the CDEs (for which assumption (i) and (ii) are sufficient enough). The biggest issue in the presence of causally dependent L is that testing the mediators one at a time will no longer be viable because this results in counting certain causal pathways more than once (VanderWeele and Vansteelandt 2014). A further problem emphasised by some (Daniel, De Stavola, and Cousens 2011) is that the direct effect of T on Y will not estimate consistently if there is an uncontrolled post-treatment confounder L that opens a backdoor path of $T \rightarrow L \rightarrow Y$. A seemingly easy fix to this problem is to condition for L as well. However, since L was affected by T controlling for L in the model for M, this blocks the $T \rightarrow L \rightarrow Y$ path, which will also result in biased estimates for the NDE and NIE[18]. This means that neither the inclusion nor the exclusion of L will solve the problem of identifiability.

Thus, to address these issues we require a *refined – and relaxed – sequential ignorability assumption* to make the NIE and NDE identifiable. Crucially, these

[18] This is the reason why in some places (e.g., Avin et al. 2005; Tchetgen Tchetgen and VanderWeele 2014) L is referred to as a "recanting witness".

alternatives do not contradict Avin et al. (2005), but introduce additional assumptions to make the natural effects estimable. While several alternative sets of identifiability criteria have been established (Tchetgen Tchetgen and VanderWeele 2014), the one postulated by De Stavola et al. (2015) will be discussed here. De Stavola et al. (2015) modified the definition of Imai and Yamamoto (2013), positing that *sequential ignorability in the presence of post-treatment confounder L* holds when controlling for pre-treatment covariates C when there is:

v. No unmeasured confounding of the T-Y, T-M, and T-L relationship or $(Y_{tml}\ M_{tl}\ L_t) \perp\!\!\!\perp T|C$
vi. No unmeasured confounding of the M-Y relationship also controlling for T and L or $Y_{tml} \perp\!\!\!\perp M|C, T, L$
vii. No unmeasured confounding of the L-Y relationship also controlling for T or $Y_{tl} \perp\!\!\!\perp L|C, T$
viii. No unmeasured M-Y confounder Z that was affected by T or $Y_{tm} \perp\!\!\!\perp M_{t*}|C, L$

These assumptions are analogous to (i-iv). *Assumption (v) is satisfied in the case of a random assignment of T*, while (vi) makes the mediated effect conditional on L. Assumption (vii) stresses that there cannot be any unmeasured confounder for the L-Y relationship, which is again a strong assumption similar to (ii). Finally, (viii) establishes that there cannot be any post-treatment confounder Z that was not included in L (i.e., all post-treatment confounders – in other words, alternative mediators – are measured).

Under assumptions (v-vii) the NDE and NIE are identifiable with a few additional limitations. Firstly, the analyst needs to rely on the *linearity assumption*, which permits the additivity of the effects. Secondly, it needs to be assumed that there is no significant T-M interaction for Y (Robins and Greenland 1992) or that both the T-L interaction and L^2 are zero in the model for Y (Petersen, Sinisi, and van der Laan 2006). Importantly, this second modelling assumption is a *loosened version of SEM's effect homogeneity assumption* that only needs to be true on average, not for each individual, thus it can be empirically assessed (Imai and Yamamoto 2013). Crucially, and as demonstrated by De Stavola et al. (2015), *when these parametric assumptions are met, (vii) is automatically satisfied*.

Provided that the model is identifiable, a generalised structural equation model needs to be specified with L modelled on C and T, M modelled on C, T, and L, and Y modelled on C, T, L, and M. In addition, the interaction between T and L is entered in the model for both M and Y, and the squared transformation of M and L in the model for Y to control for potential quadratic and heterogeneous effects in line with the earlier identification assumptions[19]. Then, a generalised version of the product method is used to obtain the parameters. G-computation of the causal estimates for NDE and NIE can be accomplished by combining these appropriate parameters from the SEM through estimation by combination. Unfortunately, with more complex models this mathematical integration can become exceedingly cumbersome with potential issues of convergence. To overcome this difficulty, Monte Carlo simulation is used as a more flexible and efficient way to approximate the integration, whilst the standard errors and confidence intervals are bootstrapped (Daniel et al. 2011). As acknowledged by De Stavola et al. (2015), the results of this approach will coincide with a traditional SEM, provided that there are no interactions or nonlinear terms of M, L, or T (for the equations discussed in this paragraph please refer to Appendix/C).

Finally, it is crucial to *assess the robustness of the M-Y relationship to potential unmeasured confounding*. De Stavola et al. (2015) devised a sensitivity analysis based on Imai et al.'s (2011) method that can be applied in the presence of post-treatment confounders. This refined method fits a SEM which allows for the error terms of the models for Y and M to become correlated[20]. These error terms are very instructive as they incorporate the impact of the unmeasured confounders. This method regresses L on X and C, M on L, X, and C, and Y on X, L, and C. M is not included in the model for Y as to do so would induce collinearity. Then the error terms from the model for M and Y are systematically correlated, where ρ' provides an indication of how big the correlation must be between the two error terms to make the M-Y relationship zero. A confidence interval for ρ' can also be obtained with bootstrapping.

[19] In particular, this is the model specification for Robins and Greenland (1992). For Petersen et al. (2006) the interaction between T and L is entered only in the model for M, and the squared transformation of M and the M-T interaction are included in the model for Y.

[20] For the causal identification of the direct and indirect effects one of the assumptions of SEM is that the errors are not correlated with each other.

Preliminary remarks

Confirmatory factor analysis was used to derive factor scores for the respective constructs in both studies. These factor scores were entered in the causal mediation analysis. Although this strategy might have resulted in increased measurement error bias compared to using latent variables to capture multiple indicators (Loeys et al. 2014), the reliance on latent variables, their interactions and transformations (i.e., squared-forms) would have added to the computational complexity and prolonged the already fairly long estimation time. Moreover, the concepts and definitions of causal effects only apply to the structural model, not the measurement models. Thus for pragmatic reasons, factor scores were used instead of latent variables to demonstrate the use of causal mediation analysis with multiple mediators.

All analyses in this paper were carried out using STATA 14 and its multicore (MP) version with g-computation models relying on 100,000 Monte Carlo simulations and the number of bootstraps set to 300. The cap on the number of bootstraps was placed so that the analysis would mirror a realistic application, as causal mediation analysis with post-treatment confounding can be particularly time-consuming. With the current specification it takes five days with a regular office computer and a single core, and two days with a cluster computer and six cores to obtain results. The estimation of causal mediation analysis with sequentially ordered mediators is speedier, taking a matter of minutes.

Test of identification

Two linear regression analyses were fitted for cooperation for the two studies. In both cases the treatment, the covariates (gender, age, ethnic minority background) and the two potential mediators, their quadratic transformation, and their interaction with the treatment were included (Table 1). As discussed earlier, including these additional parameters in the regression models for the respective Y and examining their significance helps in determining which identification strategy – if any – is relevant for a particular model. However, as noted by VanderWeele (VanderWeele 2015; VanderWeele and Knol 2014), strong statistical power is usually needed to discover interactions, therefore it is worth also examining effect sizes, especially with smaller sample sizes. In this paper, the sample sizes are moderate and the effect sizes are relatively uneqivocal, which means that considering the statistically significant effects should be sufficient.

For Study 1 either identification strategy (Petersen et al. 2006; Robins and Greenland 1992) should suffice as neither the interactions, nor the quadratic terms are significant. However, for the second study, neither will be appropriate, as there are moderately strong and significant interactions between the treatment and both mediators. The NDE and NIE can still be estimated, but they have no meaningful interpretation, thus they are not included in the results table (Table 2). Nevertheless, the TE is always estimable, as well as the CDE(m), provided that assumptions (v) and (vi) are satisfied.

	Study 1	*Study 2*
Moral alignment	0.368***	1.143***
	[0.193, 0.544]	[0.703, 1.583]
Moral alignment^2	-0.044	0.120
	[-0.144, 0.056]	[-0.018, 0.257]
Moral al. X Treatment	-0.037	-0.478***
	[-0.248, 0.174]	[-0.699, -0.257]
Duty to obey	0.193*	0.570*
	[0.014, 0.373]	[0.087, 1.107]
Duty to obey^2	-0.006	0.059
	[-0.101, 0.089]	[-0.075, 0.194]
Duty to obey X Treatment	-0.097	-0.365**
	[-0.324, 0.130]	[-0.647, -0.083]
Treatment	0.282**	1.136***
	[0.0822, 0.483]	[1.067, 1.648]
Gender	0.096	0.117
	[-0.051, 0.243]	[-0.017, 0.252]
Age (years)	-0.003	-0.006*
	[-0.051, 0.243]	[-0.012, -0.001]
Ethnic minority	-0.207*	-0.111
	[-0.374, -0.040]	[-0.300, 0.077]
Constant	0.014	-1.143**
	[-0.365, 0.392]	[-1.179, -0.497]
N	215	235

Table 1 Test of identification for post-treatment confounding, linear regression analyses with 300 bootstraps

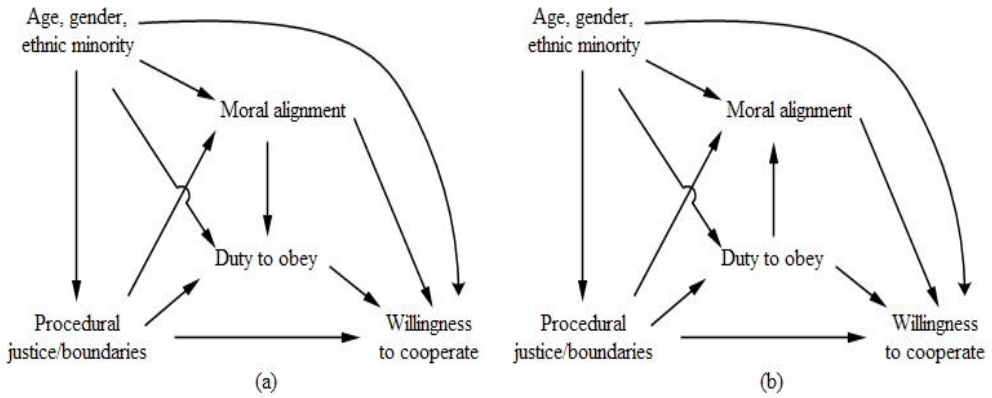


Figure 2 Mediation analysis where (a) moral alignment has a causal effect on duty to obey or (b) duty to obey has a causal effect on moral alignment

Results for causally dependent mediators

As discussed earlier, some scholars (e.g., Hough et al. 2013) believe that the two aspects of legitimacy, moral alignment and duty to obey, mutually reinforce each other. In the SEM literature this is depicted using a bidirectional arrow that denotes a correlation between the constructs. By contrast, the causal inference literature utilises directed acyclic graphs (DAGs), which do not allow two-headed arrows as to do so would create a cycle. Hence, when mutual reinforcement is hypothesised two graphs are created for the two different causal directions (Figure 2 (a) and (b)).

In causal mediation analysis with post-treatment confounding the NIE incorporates the mediated effect of the mediator of interest (including L's impact on M) and the NDE the effect of the treatment not going through M (including L's impact on Y). In Study 1 (Table 2) procedural justice treatment has a significant positive effect (NDE=0.254, $p<0.01$) and a significant mediated effect through moral alignment with the police (NIE=0.215, $p<0.001$), which carries approximately 46% of the total effect. The sensitivity analysis indicates that on average a relatively strong correlation of $\rho'=0.42$ would be needed between the error terms to nullify the mediated effect with a 95% confidence interval of 0.301 and 0.545. Conversely, duty to obey the police has a weak non-significant impact on cooperation (NIE=0.043, $p>0.05$) and transmits only 9% of the total effect. The sensitivity analysis implies that on average a $\rho'=0.199$ between the error terms could make the impact non-significant, but the confidence intervals show that a correlation of 0.047 might be enough to make the effect zero. Procedural justice has a strong and significant direct effect on willingness to cooperate (NDE=0.423, $p<0.001$). The estimates of CDE(m) and NDE are both within rounding

error of each other in both cases, which is unsurprising given the absence of a treatment-mediator interaction, in which case they should approximately coincide (i.e., as a default the CDE's m is always set at the average value of the mediator, as this option allows the comparison to the NIE). Overall, it seems that moral alignment with the police has a fairly strong causally mediated effect on cooperation with the police, while duty to obey does not seem to have an impact. Receiving the procedural justice treatment also significantly increased the participants' willingness to cooperate with the police.

	Cooperation	*Proportion mediated*	*ρ'*
Study 1 (n=215)			
Moral alignment NIE	0.215*** [0.106, 0.325]	46%	0.420 [0.301, 0.545]
Pj vs punj NDE	0.254** [0.106, 0.403]		
Pj vs punj CDE(m)	0.255** [0.107, 0.325]		
TCE	0.470*** [0.289, 650]		
Duty to obey NIE	0.043 [-0.009, 0.095]	9%	0.199 [0.047, 0.354]
Pj vs punj NDE	0.423*** [0.255, 0.591]		
Pj vs punj CDE(m)	0.423*** [0.256, 0.591]		
TCE	0.466*** [0.285, 0.647]		
Study 2 (n=235)			
Moral alignment NIE	n.i.		
Legal vs illegal NDE	n.i.		
Legal vs illegal CDE(m)	1.276*** [0.872, 1.681]		
TCE	1.755*** [1.253, 2.256]		
Duty to obey NIE	n.i.		
Legal vs illegal NDE	n.i.		
Legal vs illegal CDE(m)	1.763*** [0.123, 2.296]		
TCE	1.852*** [1.314, 2.390]		

**p<0.05, **p<0.01, ***p<0.001, n.i.=not identifiable*

Table 2 Causal mediation analysis with post-treatment confounding using Robins and Greenland's (1992) identification assumption

For Study 2, only the CDE(m) and TCE were identifiable. For both moral alignment (CDE(m)=1.276, $p<0.001$) and duty to obey (CDE(m)=1.763, $p<0.001$) as mediators the controlled direct effect of legality was significant. Unfortunately, does not permit effect decomposition, thus no information can be gained regarding the mediated effects. Still, the CDE(m) was much higher than in Study 2, which implies that legal police practices might have an even stronger effect when the police are otherwise thought to overstep legal boundaries.

Causal mediation analysis with sequentially ordered mediators

Interdependence between mediators can take the form of a causal chain where the first mediator affects the second mediator and the outcome, but not the other way around. Crucially, this situation's DAG takes the very same form as the post-treatment confounder case's as shown in Figure 1 (b) and (c) where the difference between the two graphs is only substantive. The distinction between the two approaches becomes clearer looking at Figure 1 (d) and (e), which show that in the post-treatment confounder case L_1 and L_2 do not affect one another, while in the sequential case M_1 and M_2 do, following a pre-determined order. This difference in the causal structure leads to *an alternative four-way decomposition in case of two mediators* where there will be NIE_1 standing for M_1's mediated effect on Y ($T \rightarrow M_1 \rightarrow Y$), NIE_2 for M_2's mediated effect on Y ($T \rightarrow M_2 \rightarrow Y$), NIE_{12} for M_1's and M_2's jointly mediated effect on Y ($T \rightarrow M_1 \rightarrow M_2 \rightarrow Y$), and NDE, T's effect that does not go through either of the mediators ($T \rightarrow Y$). Although there have been other approaches addressing causally ordered mediators (Steen et al. 2017a, 2017b), Daniel et al. (2015) has been the only paper so far to allow for this finest four-way decomposition. As before, this new decomposition will require a *modified set of sequential ignorability assumptions*; controlling for pre-treatment covariates C, these are:

ix. No unmeasured confounding of the T-Y, T-M_1, and T-M_2 relationship or $(Y_{tm_1m_2}\ M_{2tm_1}\ M_{1t}) \perp\!\!\!\perp T|C$

x. No unmeasured confounding of the M_1-Y relationship also controlling for T or $Y_{tm_1m_2} \perp\!\!\!\perp M_1|C,\ T$

xi. No unmeasured confounding of the M_2-Y relationship also controlling for T and M_1 or $Y_{tm_1m_2} \perp\!\!\!\perp M_2|C,\ T,\ M_1$

xii. No unmeasured M_1-Y, M_1-M_2 or M_2-Y confounder L_1 or L_2 that was affected by T or $Y_{tm_1m_2} \perp\!\!\!\perp M_{1t^*}|C$, $M_{2tm_1} \perp\!\!\!\perp M_{1t^*}|C$, and $Y_{tm_1m_2} \perp\!\!\!\perp M_{2t^{**}}|C$, M_{1t^*}

When T is randomly assigned, (ix) will automatically be satisfied. (x) is analogous to (ii) and (xi) to (vi), while (xii) states again that there cannot be post-treatment confounders L_1 or L_2 that were affected by the treatment. As with the post-treatment confounder case, certain parametric restrictions are also needed. As earlier, the *linearity assumption* is required so the additivity of the effects is guaranteed. Furthermore, when M_1 has a non-zero effect on M_2 the *conditional correlation between M_1's potential outcomes is required* to make the effects estimable. However, this conditional correlation is unknown for several of the effects[21], hence a *sensitivity parameter κ^2* is used, which stands for the proportion of residual variance shared across the two hypothetical worlds. κ^2 can take values from 0 to 1, where 0 means no correlation between the potential outcomes conditional on C, and 1 means perfect correlation between the potential outcomes conditional on C (Daniel et al. 2015).

Because of the second mediator, the conditional expectations take more complex forms: generally they are $E[Y(t,M_1(t^*),M_2(t^{**},M_1(t^{***})))]$, where the different t-s (i.e., t, t^*, t^{**}, t^{***}) stand for setting the treatment to one of its possible values. This increased complexity also means that the number of possible decompositions of the total effect will be $(2^n)!$, where n stands for the number of mediators. In the case of two mediators, the 24 (i.e., $(2^2)!=(4)!=(4x3x2x1)=24$) possible decompositions are reduced to 6 when M_1 does not affect M_2. Overall, *marked differences among the path-specific effects only emerge when there are significant T-M_1 and T-M_2 interactions, which are allowed with the current technique*. In the absence of interactions, the estimates will be approximately the same as SEM's estimates, albeit with wider confidence intervals (Daniel et al. 2015).

Because interpreting a high number of estimates from the different decompositions can be cumbersome, and usually not of particular interest, it is worth considering ways to summarise the effects. Based on earlier work (Kuha and Goldthorpe 2010), Daniel et al. (2015) recommended the usage of *summary effects that are weighted averages of the NDE and various NIEs*. In addition, they also advised

[21] When the first and the third potential outcome are set to the same value (i.e., $t^*=t^{***}$, such as NIE_1-101, NIE_1-010 etc.) this sensitivity parameter is not needed.

reporting the variance estimates for these summary effects, which indicate whether there are large differences across the various decompositions. The major advantage of these summary effects is that they provide a good approximation of the respective effects, however it is hard to attach a substantive interpretation, which can prove problematic especially if the particular effect size on the outcome variable were directly interpretable and of particular interest.

Study 1 (n=215)	*Cooperation*	*Cooperation*	*Proportion mediated*
Sensitivity parameter (κ^2)	*=0*	*=1*	*=0-1*
Moral alignment SNIE$_1$	0.231** [0.059, 0.392]	0.219** [0.063, 0.374]	42-46%
Moral alignment IE$_{1nointer}$	0.225*** [0.103, 0.347]	0.225*** [0.116, 0.333]	45%
Moral alignment $\sqrt{varNIE}_1$	0.006 [-0.020, 0.032]	0.001 [-0.007, 0.009]	
Duty to obey SNIE$_2$	-0.010 [-0.052, 0.032]	-0.009 [-0.052, 0.033]	2%
Duty to obey IE$_{2nointer}$	-0.017 [-0.050, 0.017]	-0.017 [-0.053, 0.020]	3%
Duty to obey $\sqrt{varNIE}_2$	~0.001 [-0.004, 0.004]	~0.001 [-0.005, 0.005]	
Joint SNIE$_{12}$	0.061* [0.002, 0.12]	0.038 [-0.017, 0.094]	7-12%
Joint IE$_{12nointer}$	0.064* [0.002, 0.126]	0.064* [0.007, 0.121]	12%
Joint $\sqrt{varNIE}_{12}$	0.006 [-0.017, 0.029]	0.001 [-0.006, 0.007]	
Pj vs nopj SNDE	0.218 [-0.077, 0.513]	0.273 [-0.022, 0.569]	
Pj vs nopj $\sqrt{varNDE}$	0.007 [-0.011, 0.024]	0.002 [-0.015, 0.018]	
TCE	0.499** [0.166, 0.834]	0.521** [0.180, 0.863]	

**p<0.05, **p<0.01, ***p<0.001, pj=procedural justice, nopj=procedural injustice*

Table 3 Causal mediation analysis with sequentially ordered mediators, Study 1

Results for sequentially ordered mediators

In the procedural justice literature, other scholars (e.g., Huq, Jackson, and Trinkner 2017) have argued for the theoretical model depicted by Figure 2 (b), where duty to obey the police is influenced by moral alignment with the police, but not the other way around. The results (Tables 3-4) are presented conditionally only on the two extremes

of the sensitivity value (κ^2=0 and κ^2=1), as they appear to be mostly robust to these extremes. In the case of wider disparities, it is advisable to look at further values of the sensitivity parameter.

The results from Study 1 (Table 3, Appendix Figure 1/a-4/a) show that moral alignment with the police has a moderately strong mediated effect ($SNIE_1^{\kappa2=0}$=0.231, p<0.01, $SNIE_1^{\kappa2=1}$=0.219, p<0.01) on willingness to cooperate with the police with 42-46% of the effect transmitted by it. In contrast, duty to obey does not have an impact on any conventional level of statistical significance ($SNIE_2^{\kappa2=0}$=-0.01, p>0.05, $SNIE_2^{\kappa2=1}$=-0.009, p>0.05). The joint effect of moral alignment with and duty to obey the police has a weak significant relationship with willingness to cooperate when κ^2=0 ($SNIE_{12}^{\kappa2=0}$=0.061, p<0.05), but it does not reach the 5% significance level when κ^2=1 ($SNIE_{12}^{\kappa2=1}$=0.038 p>0.05). Procedural justice treatment has a moderately strong effect on cooperation, but it does not reach statistical significance ($SNDE^{\kappa2=0}$=0.218, p>0.05, $SNDE^{\kappa2=1}$=0.273, p>0.05). Juxtaposing the results of the two extremes of the sensitivity parameter shows that when the counterfactual outcomes of M_1 are assumed to be perfectly correlated, it slightly boosts the SNDE and widens its confidence intervals but at the same time reduces the mediated effects and narrows their confidence intervals. The variance estimates are tiny across the different effects, which is in accordance with the lack of interactions found earlier (Table 1). The absence of interactions also means that the indirect effects (IEs) from an SEM should be very close to the SNIEs which, as expected, they are.

The findings from Study 1 seem to uphold moral alignment's moderately strong mediated effect on willingness to cooperate while also confirming the lack of impact from duty to obey. The joint effect of the two mediators is either weak or non-existent while the direct effect of procedural justice does not reach statistical significance; a bigger sample size would be needed to elucidate the treatment's and the joint effect's impact on the outcome.

The results from Study 2 (Table 4, Appendix Figure 5/a-8/a) show a similar pattern to Study 1. Moral alignment with the police has a strong mediated effect ($SNIE_1^{\kappa2=0}$=0.623, p<0.001, $SNIE_1^{\kappa2=1}$=0.640, p<0.001) with 32% proportion mediated while duty to obey has a weak non-significant one ($SNIE_2^{\kappa2=0}$=0.081, p>0.05, $SNIE_2^{\kappa2=1}$=0.081, p>0.05). Yet again the joint mediated effect is either significant ($SNIE_{12}^{\kappa2=1}$=0.184, p<0.05) or not ($SNIE_{12}^{\kappa2=0}$=0.178, p>0.05), depending on the value taken by the sensitivity parameter. The direct effect of legality is much

stronger in Study 2 than procedural justice's in Study 1 with a statistically significant impact ($SNDE^{\kappa^2=0}$=1.085, p<0.001, $SNDE^{\kappa^2=1}$=1.085, p<0.001). The results are even less sensitive to changes in κ^2 than in Study 1, and the perfect correlation between the potential outcomes of M_1 here increases the effect sizes while decreasing the confidence intervals for all effects. The variance estimates are much higher than in Study 1, which is expected because of the interaction effects (Table 1). This also means that indirect effects conventionally obtained from an SEM will be different: in this case they are much smaller than the ones from causal mediation analysis.

Study 2 (n=235)	*Cooperation*	*Cooperation*	*Proportion mediated*
Sensitivity parameter (κ^2)	*=0*	*=1*	*=0-1*
Moral alignment $SNIE_1$	0.623*** [0.321, 0.926]	0.640*** [0.358, 0.921]	32%
Moral alignment $IE_{1nointer}$	0.424*** [0.240, 0.609]	0.441*** [0.277, 0.605]	27%
Moral alignment $\sqrt{varNIE}_1$	0.0.22 [-0.032, 0.077]	0.023 [-0.010, 0.057]	
Duty to obey $SNIE_2$	0.081 [-0.045, 0.206]	0.081 [-0.032, 0.195]	5%
Duty to obey $IE_{2nointer}$	0.023 [-0.016, 0.062]	0.023 [-0.015, 0.061]	1%
Duty to obey $\sqrt{varNIE}_2$	0.001 [-0.005, 0.007]	0.001 [-0.003, 0.007]	
Joint $SNIE_{12}$	0.178 [-0.015, 0.370]	0.184* [0.033, 0.336]	9%
Joint $IE_{12nointer}$	0.060 [-0.036, 0.123]	0.062* [0.001, 0.124]	4%
Joint $\sqrt{varNIE}_{12}$	0.006 [-0.036, 0.047]	0.005 [-0.005, 0.015]	
Pj vs illegal SNDE	1.085*** [0.749, 1.421]	1.085*** [0.737, 1.432]	
Pj vs illegal $\sqrt{varNDE}$	0.039 [-0.007, 0.085]	0.041 [-0.001, 0.083]	
TCE	1.967*** [1.364, 2.569]	1.990*** [1.424, 2.556]	

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

Table 4 Causal mediation analysis with sequentially ordered mediators, Study 2

Overall the results provide further support for the earlier findings. From the two components of legitimacy shared values (moral alignment) appears to be the important mediator of legality's impact on cooperation while consent (duty to obey)

does not seem to matter much. The joint effect of these two elements is very close to zero and requires further scrutiny. The strong direct effect indicates that procedurally just and legal messaging will have a powerful impact, especially when compared to the assumption that the police routinely overstep their boundaries.

In summary, both the post-treatment confounder and sequentially ordered approach concur that *moral alignment is the primary conduit of the effect of procedural justice and legality on willingness to cooperate, while duty to obey either does not have an effect or has only a weak joint one with moral alignment*. The direct effect of the two treatment conditions also seemed to be important, even if not consistently significant between the two methods. These results are in line with earlier research (e.g., Moravcová 2016; Tyler and Jackson 2014), which found a small or even non-significant relationship between duty to obey and cooperation, and which thus called into question its relevance. As with other experiments, the external validity of the results is limited and further studies are needed to attest to the effects found here.

Discussion

Over the last couple of decades, many social science disciplines have relied primarily on SEM (and path analytical models more generally) for assessing complex theories. Yet, adopting the potential outcome framework provides at least three advantages (Daniel, Stavola, and Vansteelandt 2016; Greenland 2017; Steen et al. 2017a):

- it makes explicit the identification assumptions needed to avoid model misspecification for the mediator(s);
- it provides formal definitions of the estimated causal effects; and,
- it devises ways to check for the robustness of the results through sensitivity analysis of certain causal identification assumptions[22].

This paper has argued that the traditional SEM framework has shortcomings that need to be addressed for more realistic identification and effect decomposition. In order to accommodate multiple mediated effects, parametric restrictions akin to SEM

[22] Importantly, sensitivity analyses are not exclusive to the potential outcome framework, SEM has been also applying such techniques typically to test certain modelling assumptions (e.g., Pek and MacCallum 2011), but sometimes also to assess causal identifying assumption (Mauro 1990).

Mediation analysis technique	*Causal and parametric assumptions*
Single mediator *Causal independence (Figure 1 (a))* *Single vector of mediators (assumptions for a vector of mediators)*	i. No unmeasured confounding of the T-Y relationship or $Y_{tm} \perp\!\!\!\perp T|C$ ii. No unmeasured confounding of the M-Y relationship also given T or $Y_{tm} \perp\!\!\!\perp M|C,T$ iii. No unmeasured confounding of the T-M relationship or $M_t \perp\!\!\!\perp T|C$ iv. No unmeasured M-Y confounder L that was affected by T or $Y_{tm} \perp\!\!\!\perp M_{t^*}|C$ Non-parametrically identifiable.
Post-treatment confounding (Figure 1 (b) and (d))	v. No unmeasured confounding of the T-Y, T-M, and T-L relationship or $(Y_{tml}\ M_{tl}\ L_t) \perp\!\!\!\perp T|C$ vi. No unmeasured confounding of the M-Y relationship also controlling for T and L or $Y_{tml} \perp\!\!\!\perp M|C, T, L$ vii. No unmeasured confounding of the L-Y relationship also controlling for T or $Y_{tl} \perp\!\!\!\perp L|C, T$ viii. No unmeasured M-Y confounder Z that was affected by T or $Y_{tm} \perp\!\!\!\perp M_{t^*}|C, L$ Linearity and some kind of no-interaction.
Sequential order (Figure 1 (c) and (e))	ix. No unmeasured confounding of the T-Y, T-M_1, and T-M_2 relationship or $(Y_{tm1m2}\ M_{2tm1}\ M_{1t}) \perp\!\!\!\perp T|C$ x. No unmeasured confounding of the M_1-Y relationship also controlling for T or $Y_{tm1m2} \perp\!\!\!\perp M_1|C, T$ xi. No unmeasured confounding of the M_2-Y relationship also controlling for T and M_1 or $Y_{tm1m2} \perp\!\!\!\perp M_2|C, T, M_1$ xii. No unmeasured M_1-Y, M_1-M_2 or M_2-Y confounder L_1 or L_2 that was affected by T or $Y_{tm1m2} \perp\!\!\!\perp M_{1t^*}|C$, $M_{2tm1} \perp\!\!\!\perp M_{1t^*}|C$, and $Y_{tm1m2} \perp\!\!\!\perp M_{2t^{**}}|C$, M_{1t^*} Linearity and influence of sensitivity parameter κ^2 (when M_1 affects M_2).

Table 5 Summary of the causal and parametric assumptions of the causal mediation analysis techniques discussed in the paper

need to be made: linearity and relaxed effect homogeneity for the post-treatment confounder, and linearity for the causally ordered case (for a summary see Table 5). The similarity between the two approaches does not end there; traditional SEM can be considered a special case of causal mediation analysis when certain conditions apply. This means that SEM and causal mediation analysis can be easily reconciled, and that the estimation method will be very similar to each other, which makes such techniques easily understandable and adaptable for those who were primarily trained for SEM (Daniel et al. 2015; De Stavola et al. 2015).

Study 1 and Study 2 exemplify how SEM compares to causal mediation analysis with multiple mediators. The results from Study 1 were approximately identical to the results one would have derived using SEM. In contrast, for Study 2 the post-treatment confounder case was not identifiable, while the sequentially ordered mediator case differed decidedly from the SEM results. Study 1 highlights how traditional SEM can sometimes hit the mark, while Study 2 illustrates that it can also fail. The sensitivity analysis from the different studies can help to determine whether the results are robust to certain conditions (unmeasured confounding or the correlation of certain potential outcomes). The reliance on these sensitivity measures can mitigate bad practices, like "p-hacking" and can help to identify spurious relationships and statistical flukes. This perspective also encourages researchers to adapt a priori model building since their decision will have a major impact on the modelling strategy employed, and because the causal structure can never be decided by relying on statistical methods.

Nevertheless, there are certain limitations worthy of discussion. First, causal mediation analysis relies on very strong assumptions. Even in the case of a randomly assigned treatment, the M-Y relationship can be spurious unless a proper set of covariates is controlled for. In Study 1 and Study 2 only three covariates (age, gender, and ethnic minority background) were considered, which are far from being sufficient (Steiner et al. 2010). The problem of no unmeasured confounding is further aggravated in observational studies where the treatment is not randomised. Some scholars have recommended conducting 'comprehensive SEM' (Mackinnon and Pirlott 2015) with up to fifty covariates, yet even in such cases it can be difficult to realistically argue for causal inference. Multiple mediators can even exacerbate this issue as it is more likely that at least one of them is affected by unmeasured confounding (VanderWeele 2015).

As an alternative to natural effects, some have recommended the use of interventional effects (Vansteelandt and Daniel 2017), which require weaker causal identifying assumptions. However, these loosened assumptions posit additional parametric restrictions to the ones that have been discussed in this paper (e.g., fixing the mediator distribution). Arguably, these alternatives can sometimes be more policy relevant (i.e., the interventional indirect effects are set at the levels of the potential interventions), but they provide less information regarding the causal mechanisms and hence are often times less generalisable to other contexts.

Others are also critical of SEMs because of their restrictive parametric assumptions, which were (partially) adopted for causal mediation analysis in the current applications. VanderWeele has repeatedly insisted (VanderWeele 2012, 2015, 2016) that these modelling assumptions are too strong, and SEMs and similar methods should only be used for hypothesis generation not hypothesis testing. In addition, both Keele (2015a) and Kennedy (2015) have argued that because causal effects are non-parametrically identified, parametric models are more likely to yield misspecification and the use of semi- or non-parametric models is more advisable. Although such alternatives are available for multiple mediators (Kim, Daniels, and Hogan 2017; Moerkerke, Loeys, and Vansteelandt 2015; Tchetgen Tchetgen and VanderWeele 2014), they have other restrictions and limitations (e.g., particular types of outcome, constrained effect decomposition, Bayesian model-specification) that make them unappealing or hard to implement.

Even if these criticisms are valid, most of the propositions made in this paper touch upon the fundamental limitations of SEM and can be considered as improvements upon it. For instance, the no-interaction assumption is a non-causal issue, yet applying a causal mediation perspective helps to address the matter. Similarly, the current methods allow to incorporate quadratic terms in the model, provide an alternative way to investigate cases when mediators are assumed to mutually reinforce each other, and propose sensitivity analysis for model assessment. In the end, causal mediation analysis provides a list to consider for causal analysis, a slightly modified estimation approach that allows for a more versatile model analysis and assessment, and provides a comprehensive improvement upon the traditional SEM.

Appendix/A – Detailed Overview of the studies

Study 1 and Study 2 were conducted in July 2013 in two subsequent weeks on the Amazon Mechanical Turk website. Study 1 manipulated police procedural justice, while Study 2 manipulated police legality. These studies used a very similar newspaper article about road side checks in the United States as manipulation. In Study 1, the text described a procedurally unjust roadside check (i.e., angry, unresponsive, yelling officers), which was later either bolstered by fictitious data as an ordinary case (procedurally unjust condition) or as something which was an exception from the rule (procedurally just case). Study 2 introduced an almost identical story where during the roadside check the officers clearly abused their power (i.e., through excessive use of force, handcuffing and flooring an innocent driver), which was later presented either as a usual occurrence (illegal condition) or an increasingly unlikely one (legal condition). Procedural justice of the police, legality of the police, normative alignment with the police, obligation to obey the police, and willingness to cooperate with the police were measured with three items each on a 1-5 Likert-scale almost exclusively with construct-specific response alternatives (for the questionnaire and the prompts please refer to Table 1a-Table 3a).

Construct	*Questions*
Procedural justice	*Now some questions about when the police deal with crimes like house burglary and physical assault. (Almost never, Not very often, Often, Very often, Almost all the time)* Based on what you have heard or your own experience how often would you say the police generally treat people in the United States with respect. About how often would you say that the police make fair, impartial decisions in the cases they deal with? When dealing with people in the United States, how often would you say the police generally explain their decisions and actions when asked to do so? About how often would you say that the police provide opportunity for unfair decisions to be corrected?
Legality/ Boundaries	*Do you agree or disagree with each of the following statements about the police in your community? (Strongly disagree – Strongly agree)* The decisions and actions of the police are unduly influenced by pressure from political parties and politicians. The police take bribes. The police often arrest people for no good reason.
Moral alignment	*Do you agree or disagree with each of the following statements about the police in your community? (Strongly disagree – Strongly agree)* The police generally have the same sense of right and wrong as I do.

with the police	The police usually act in ways consistent with your own ideas about what is right and wrong. The police stand up for moral values that are important to people like me.
Duty to obey the police	*Do you agree or disagree with each of the following statements about the police in your community? (Strongly disagree – Strongly agree)* You should do what the police tell you to do even when you do not like the way they treat you. You should accept the decisions made by police, even if you think they are wrong. You should do what the police tell you to do even if you disagree.
Willingness to cooperate with the police	*If the situation arose, how likely would you be to: (Very unlikely – Very Likely)* …call the police to report a crime you witnessed? …report suspicious activity near your house to the police? … provide information to the police to help find a suspected criminal?
Demographic questions	Gender (*Male, Female*) How old are you? (*Free space given to fill it out*) How do you describe yourself? (*American Indian or Alaska Native; Hawaiian or Other Pacific Islander; Asian or Asian American; Black or African American; Hispanic or Latino; White*) Which state do you live in? (*state names*)

Table 1/a Questionnaire used for Study 1 and Study 2

Police rudeness and roadside checks in [State name] –

Isolated cases or business as usual?

On the night of June 22nd around 11pm, Michael Harrison was driving on a highway next to [second largest city in the State]. He was coming back from a visit to his sister and her new-born daughter, which ran a little late. Mr Harrison was listening to the radio when he was suddenly stopped by two police officers. He was not surprised, as there had been rumours of police checks in the area. He slowly pulled over his car and leaned over to the glove compartment to get his driver's license, when one of the officers started yelling at him ordering him to leave his hands on the wheel without doing any sudden movements. After telling the officers that he had been only heading home he was asked to get out of the car. Still shouting, one of the officers ordered Mr Harrison to put his hands on the engine hood then strip searched him, presumably looking for weapons. When Mr Harrison asked what he had done, one of them told him that they would let him know later. Meanwhile, the other policeman was looking inside and under the car, finally asking what he was keeping in the trunk. Mr Harrison confessed that he kept only some tools in there and was allowed to show the officers the trunk. The policemen appeared to be

really dissatisfied with the results. They took an alcohol and drug test but Mr Harrison tested negative for both. They also checked his driver's license's validity. After finding everything in order, they told him that he could leave, without giving any further explanation for the purpose of the search. Feeling humiliated, Michael Harrison drove home and called his sister to tell her what happened. Together they decided to contact the press in the morning instead of the authorities. "They made me feel ashamed" told Mr Harrison to our reporter. "I couldn't let them get away with this."

Procedurally just

In **contrast** with Michael Harrison's case, figures recently released by the FBI indicated that in [State name], complaints regarding police behaviour during roadside checks have sharply **decreased** from **201** in 2007 to an all-time **low** figure of **175** in 2012. "We are aware of the problem" admitted the police chief of [second biggest city in the State]. "that's why we try to enrol as many police officers to the training programmes as possible. I am sure that such efforts will pay off eventually."

Procedurally unjust

In **line** with Michael Harrison's case, figures recently released by the FBI indicated that in [State name], complaints regarding police mistreatment during roadside checks have sharply **increased** from **175** in 2007 to an all-time **high** figure of **201** in 2012. "We are aware of the problem" admitted the police chief of [second biggest city in the State]. "that's why we try to enrol as many police officers to the training programmes as possible. I am sure that such efforts will pay off eventually."

Table 2/a Manipulation text for Study 1

Brutal stop and search in [State name] –

Isolated cases or business as usual?

On the night of June 22nd around 11pm, Michael Harrison was driving on a highway next to [second largest city in the State]. He was coming back from a visit to his sister and her new-born daughter, which ran a little late. Mr Harrison was listening to the radio when he was suddenly stopped by two police officers. He was not surprised, as there had been rumours of police checks in the area. He slowly

pulled over his car and leaned over to the glove compartment to get his driver's license, when one of the officers started yelling at him and pointing a gun straight at him. He was ordered to get out of the car with his hands on the back of his head, he floored and handcuffed as soon as he obeyed. While the policemen searched his car he was left in the dust with his face down, asking for explanations to no vail. After they finished searching, the policemen got him up, and asked him what his purpose of being there was. He told them he had just been heading home. The officers then informed him that they had to treat him that way, because he made a threatening move and they suspected he was hiding a gun somewhere in the car. Finally, they uncuffed him and let him go after warning him, not to provoke such measures again. Still terrified, Michael Harrison drove home and called his sister to tell her what happened. Together they decided to contact the press in the morning instead of the authorities. They were afraid of possible retaliation by the local police. "I still cannot be sure that they won't come for me tomorrow or the day after that" told Mr Harrison to our reporter. "But I couldn't let them get away with this."

Legal

In **contrast** with Michael Harrison's case, figures recently released by the FBI indicated that in [State name], complaints regarding police mistreatment during roadside checks have sharply **decreased** from **201** in 2007 to an all-time **low** figure of **175** in 2012. The report also found that the various police forces conducted **impartial and thorough** internal investigations of such complaints, followed by **harsh sanctions** against convicted officers. "I am not sure whether they broke any rules" added the police chief of [second biggest city in the State]. "At such a late hour, at the side of a highway plunged in almost complete darkness, they [the officers] had to decide really quickly how to react. We will certainly look into the case, but I cannot tell right now whether they did the right thing or not."

Illegal

In **line** with Michael Harrison's case, figures recently released by the FBI indicated that in [State name], complaints regarding police mistreatment during roadside checks have sharply **increased** from **175** in 2007 to an all-time **high** figure of **201** in 2012. The report also found that the various police forces conducted

partial and sloppy internal investigations of such complaints, followed by **limited sanctions** against convicted officers. “I am not sure whether they broke any rules” added the police chief of [second biggest city in the State]. “At such a late hour, at the side of a highway plunged in almost complete darkness, they [the officers] had to decide really quickly how to react. We will certainly look into the case, but I cannot tell right now whether they did the right thing or not.”

Table 3/a Manipulation text for Study 2

Sampling

For both studies the respondents were recruited using Amazon's Mechanical Turk (AMT). AMT provides an online marketplace where Mechanical Turk Workers (or Turkers) solve Human Intelligence Tasks (HITs) uploaded by the respective Mechanical Turk Requesters (or Providers). AMT provides more diversity than a regular college sample or even an average internet survey would do (Buhrmester, Kwang, and Gosling 2011). Comparison of studies ran online and conducted in real world settings showed very similar results, which implies the transferability of the experiments (Chandler, Mueller, and Paolacci 2014; Horton, Rand, and Zeckhauser 2011; Paolacci, Chandler, and Ipeirotis 2010). In another study (Buhrmester et al. 2011) most of the data reached the required psychometric standards indicating similar internal consistencies as traditional samples. In addition, Turkers seem to be more attentive to the tasks on hand and therefore more susceptible to experimental manipulation (Hauser and Schwarz 2016). AMT has also been proven to be the most cost-effective compared to other online convenience samples (Antoun et al. 2016).

In all three studies no restrictions were made for the Turkers' characteristics other than their geographical location (i.e., United States). Although some (Peer, Vosgerau, and Acquisti 2013) suggested relying on experienced participants with proven track records who were less likely to fail the attention tests (thus, providing fewer exclusions), this would have increased the selection bias, hence this filter was not imposed. In line with Mason and Suri's (2012) recommendation, the two biggest Turker sites, "Turkopticon" and "Turker nation" were monitored during the data collection period to ensure that the stable unit treatment value assumption (SUTVA) was not violated. For the two studies an average of two forum entries were made, and the majority of them encouraged other fellow Turkers to fill out the surveys as they were considered a "good deal", "fascinating", "enticing" and so on. Nevertheless, none of these comments provided any information regarding the content of the studies other than the subject matter (i.e., police related survey).

Procedure

For the three surveys the Qualtrics website was used. In the beginning of the questionnaire, instructional manipulation checks were used asking the respondents to skip answering one of the questions. Those people who were inattentive enough to choose an answer alternative were eliminated from the study. As noted by

Oppenheimer et al. (2009) instructional manipulation checks are not only useful because they are able to unveil impetuous satisfiers but they also encourage the individual to remain focused as further checks might show up later in the questionnaire. As discussed earlier, all of the experiments relied on textual priming as manipulation. Prior to being exposed to the manipulation respondents were reminded to read them carefully since questions might be asked regarding the content of the upcoming text. The prompts were tailored so each respondent would receive a story situated in her respective state's second largest city. This personalisation was designed to augment the story's psychological proximity for the respondents and enhance their personal involvement (Maglio, Trope, and Liberman 2013). In other words, this state-specific manipulation meant to improve the strength of the priming through the immediacy of the described situation. Following the treatment appropriate attention checks ("screeners"), questions were asked from the participants to prove that they had actually read the piece (Berinsky, Margolis, and Sances 2016). In both studies those who failed to answer correctly at least one of the questions were presumed to be running through the survey and were excluded from further analysis (but not the study itself).

All batches of questions were presented in separate blocks and in each block their order of appearance was randomised, which aimed to attenuate the potential primacy effect of the questions (Malhotra 2008). Some studies indicated that item placement can have a slight impact on the answers of the participants, which was also addressed through this randomisation (Tourangeau, Couper, and Conrad 2013). At the end of the questionnaire, participants were debriefed about the purpose of the study, given the option to share their thoughts regarding the questionnaire, and offered the opportunity to withdraw their answers from the study without forfeiting their reward. All studies went through thorough ethical consideration and received departmental approval.

Manipulation checks

For each study manipulation checks were conducted (Mutz and Pemantle 2016). For both studies these manipulation checks revealed that the procedural justice and legality conditions had the expected impact. To aid the interpretation of the results for the scales their means were derived.

Study 1:

After filtering out the participants who failed the attention checks (6 people) or decided to withdraw from the study (5 people) approximately the same number of respondents remained in each group (procedurally unjust=113, procedurally just=112). Taking the mean of the variables, the pre-treatment variables still did not show any significant difference (t_{gender}=1.17, p>0.05; t_{age}=0.93, p>0.05; t_{ethnic}=0.61, p>0.05). In contrast, procedural justice (t_{pjust}=5.98, p<0.001, M_{punj}=2.46, M_{pjust}=3.08) and legality (t_{legal}=5.09, p<0.001, M_{punj}=2.66, M_{pjust}=3.22) varied according to the procedural justice manipulation with higher values for the procedurally just, and lower values for the procedurally unjust experimental conditions.

Study 2:

Study 2 relied on an experimental design similar to Study 1, but instead of procedural justice, here, legality was manipulated. Despite the filtering for the failed attention checks (9 people) and withdrawals from the study (6 people), nearly the same number of respondents entered each experimental group (illegal=117, legal=118). The pre-treatment variables appeared to be balanced (t_{ethnic}=-1.78, p>0.05; t_{age}=1.07, p>0.05; t_{gender}=1.57, p>0.05). Legality (t_{legal}=8.91, p<0.001, M_{illeg}=2.62, M_{leg}=3.49) and procedural justice (t_{pjust}=5.26, p<0.001, M_{illeg}=2.60, M_{leg}=3.12) all exhibited significantly higher values in the legal condition than in the illegal one.

Appendix/B – Causal mediation analysis with g-computation

G-computation was first introduced by Robins (1987) as an estimation method for the causal effects of time-varying treatment in the presence of time-varying confounders which were affected by the treatment. This has been a widely applied method in epidemiology for estimating various kinds of causal effects and as an adjustment technique to derive population average (marginal) effects (Kang et al. 2014; Snowden, Rose, and Mortimer 2011; Vansteelandt and Keiding 2011; Wang, Nianogo, and Arah 2017). Regardless of any specific application, g-computation requires very similar procedures, thus for the sake of simplicity a single mediator application will be reviewed (for further details see: Wang and Arah 2015). Generally speaking, g-computation takes the following four steps:

a) *Deriving the empirical parameters where mediator M is modelled over treatment T and covariates C, and outcome Y is modelled over M, T, and C.* This model is the same as the conventionally specified model for indirect effects (which can now include a treatment-mediator interaction) and is sometimes referred to as "Q-model".
b) *Simulating the potential outcomes for the mediator and outcome relying on (a).* Simulations often take a Monte Carlo approach where the goal of the simulation is to provide a full dataset with counterfactual outcomes that are free of confounding under the causal assumptions of the sequential ignorability assumption. First, this simulation creates a sufficiently large number of copies of the original sample with C that are marginally independent of each other and T. Then M is simulated as a function of these marginally independent T and C using the parametric model obtained at (a). Finally, Y is simulated as a function of the simulated M, T, and C using the parametric model obtained at (a). This step is called g-computation.
c) *Fitting the final models on the simulated dataset of (b).* The simulated dataset from (b) is utilised to regress each different Y on T to acquire the point estimates of the marginal effects. In case of causal mediation, this will be a Marginal Structural Model.
d) *Obtaining standard errors and confidence intervals.* The default standard errors and confidence intervals generated by software programs are usually

inappropriate for g-computation parameters, which will require a resampling-based methodology such as bootstrapping.

Parametric g-computation has several advantages. It can derive various types of estimates, incorporate nonlinearites, and address different types of outcomes. Moreover, it accommodates interaction effects for both the treatment and mediator while still permitting the estimation of a single marginal effect (Daniel, De Stavola, and Cousens 2011; Wang and Arah 2015). Furthermore, the estimates of g-computation tend to yield greater robustness, stability, and precision, than the ones acquired through inverse probability weighting, especially for continuous variables (Moerkerke et al. 2015). However, and importantly, g-computation does not differ from a more conventional mediation analysis in that a misspecified model at step (a) will lead to biased estimates. Specifically, for mediation analysis this requires correct model specification for both the mediator and the outcome.

Appendix/C – The equations and assumptions needed for parametric identification

For the post-treatment confounder case three equations need to be specified (De Stavola et al. 2015): one for the mediator of interest (M), one for the outcome (Y), and one for the post-treatment confounder (L). In each of these, T stands for the treatment, C for a vector of pre-treatment covariates, M for the mediator, and L for the post-treatment confounder. The subscripts for each coefficient also indicate its connection to the particular variable. In addition, subscript 0 refers to the intercept and ε comprises the residuals for the particular equation. Thus, and as described in the main text of the article, the following general model permits the derivation of the NIE and NDE, provided that the causal identification and parametric assumptions are met (also see the same formulation in De Stavola et al. 2015: 68pp.):

$$1(a) \quad \begin{cases} L=\gamma_0+\gamma_x X+\gamma_c C+\varepsilon_l \\ M=\alpha_0+\alpha_t T+\alpha_l L+\alpha_c C+\alpha_{cl} CL+\varepsilon_m \\ Y=\beta_0+\beta_t T+\beta_l L+\beta_{ll} L^2+\beta_m M+\beta_{mm} M^2+\beta_c C+\beta_{tl} TL+\beta_{tm} TM+\varepsilon_y \end{cases}$$

As mentioned in the main text, there are two alternative ways for the parametric identification of the natural effects. For the first solution, following Robins and Greenland (1992), the interaction effect between the mediator and the treatment (β_{tm}) has to be zero. Alternatively, and as shown by Petersen et al. (2006), both the interaction effect between the post-treatment confounder and the treatment (β_{tl}) and the effect of the squared transformation of the post-treatment confounder (β_{ll}) must be zero. Considering all these, the final equation for the outcome Y either contains β_{tl} and β_{ll} (Robins and Greenland, i.e., $\beta_{tm}=0$) or β_{tm} (Petersen et al., i.e., $\beta_{tl}=\beta_{ll}=0$). In addition, and as discussed elsewhere, the linearity assumption also needs to be maintained to make the NDE and NIE estimable. Finally, the chosen model goes through the steps of g-computation (as exemplified in Appendix/B).

Figures for the Appendix

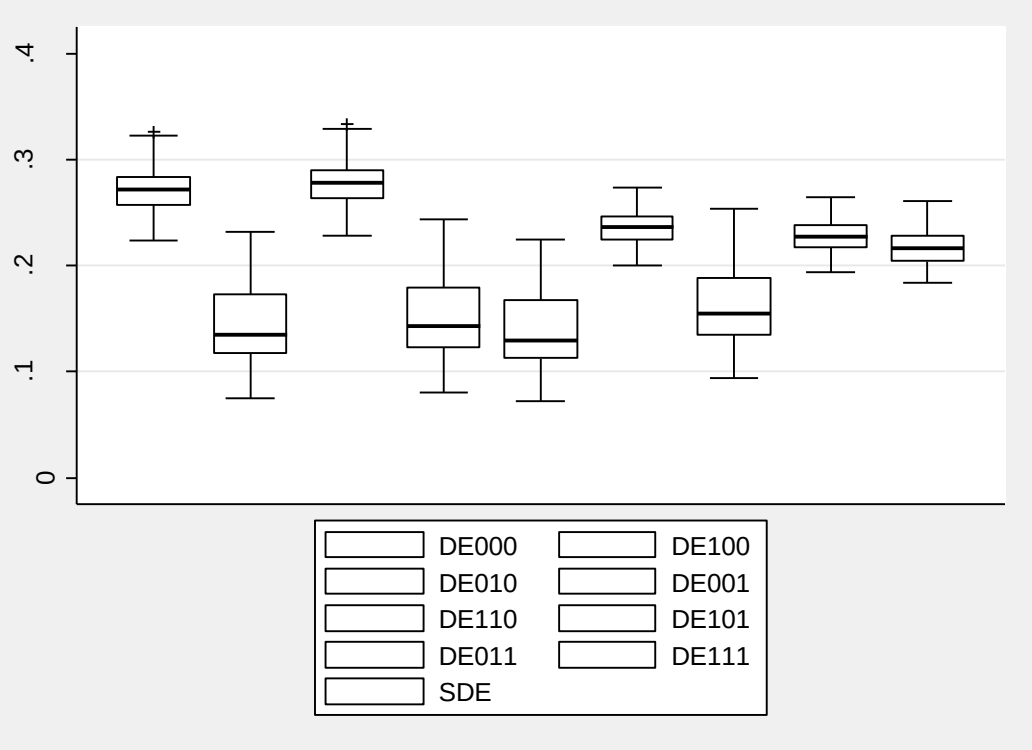


Figure 1/a NDE Procedural justice – Study 1

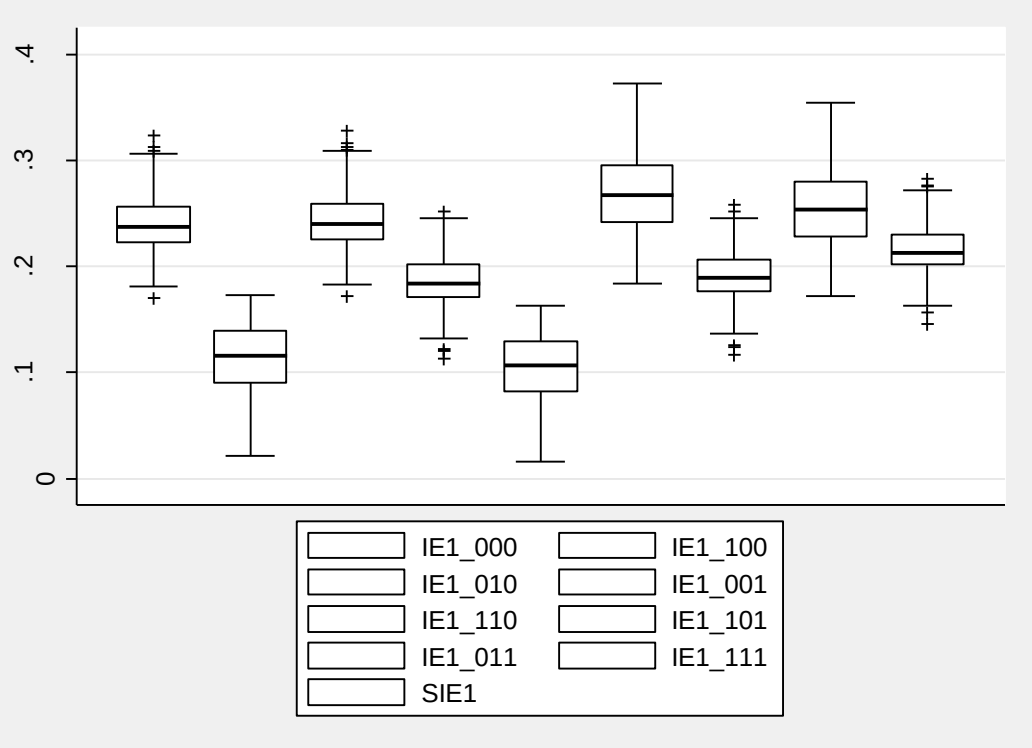


Figure 2/a NIE_1 Moral alignment – Study 1

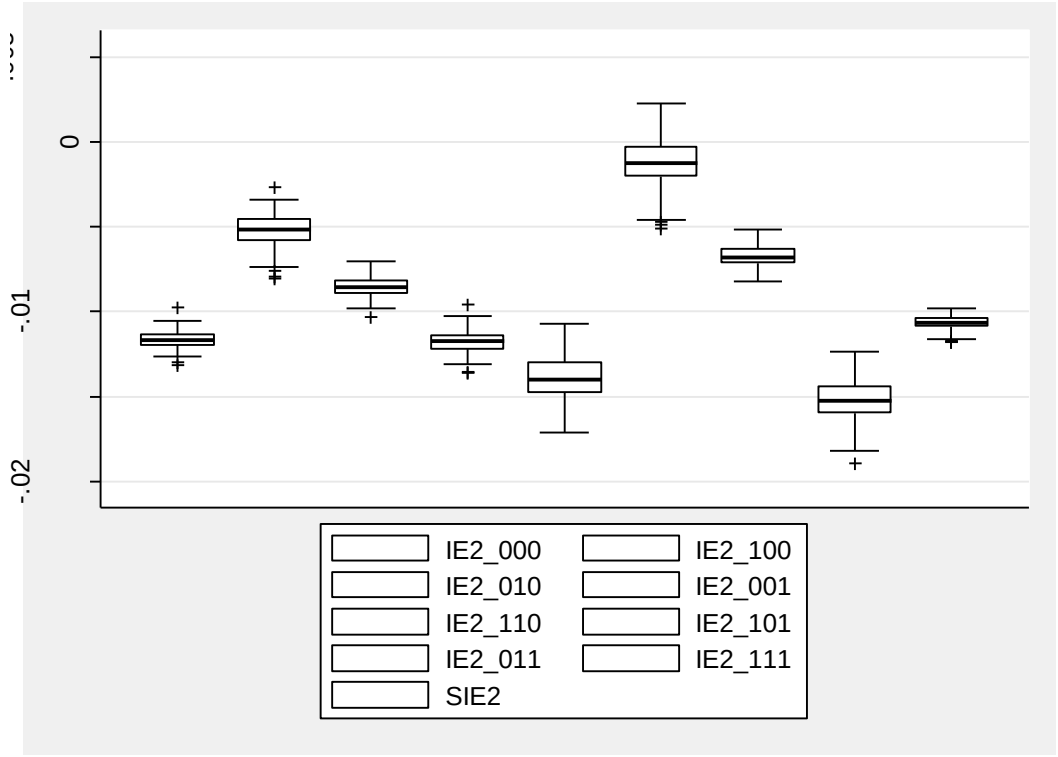


Figure 3/a NIE_2 Duty to obey – Study 1

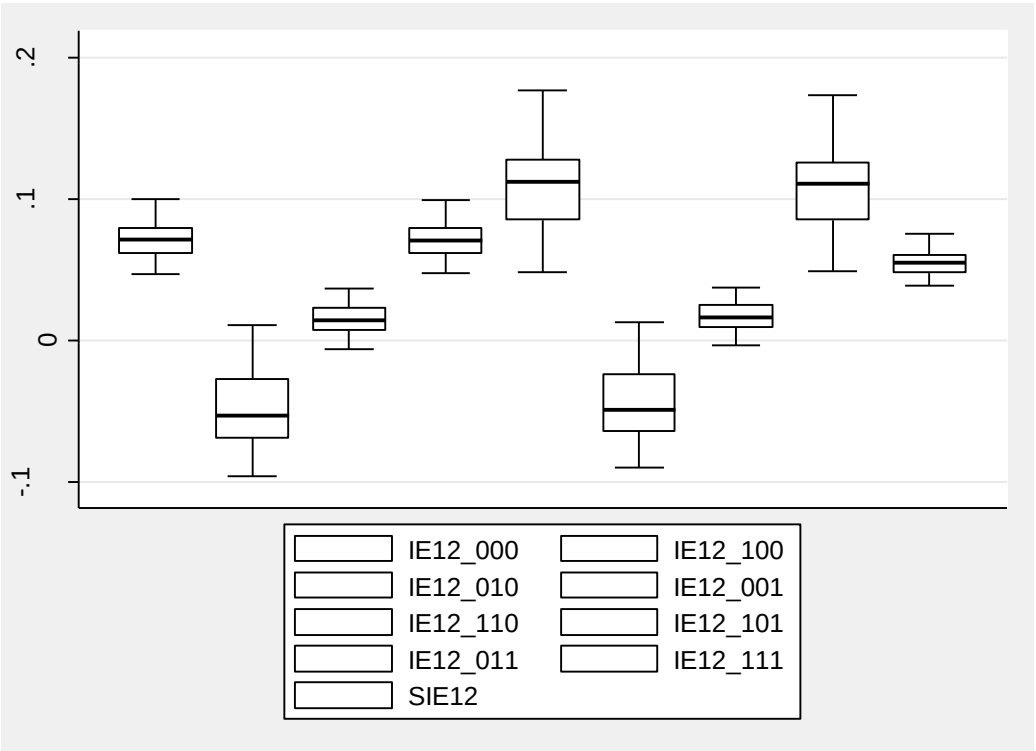


Figure 4/a NIE_{12} Joint effect – Study 1

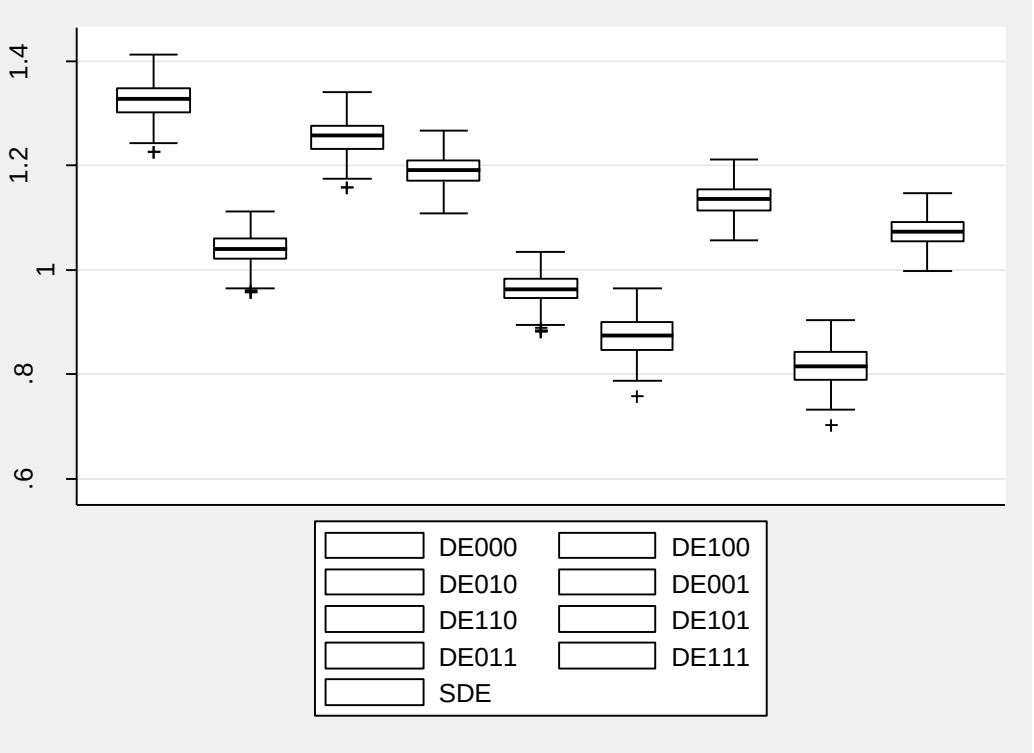


Figure 5/a NDE Legality – Study 2

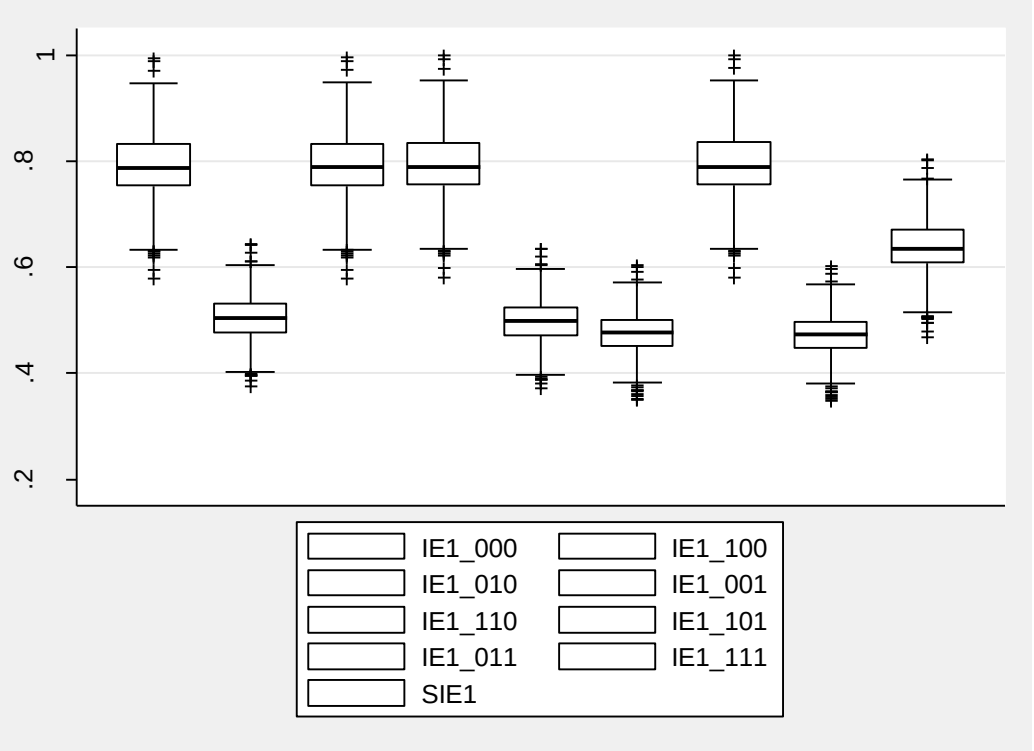


Figure 6/a NIE$_1$ Moral alignment – Study 2

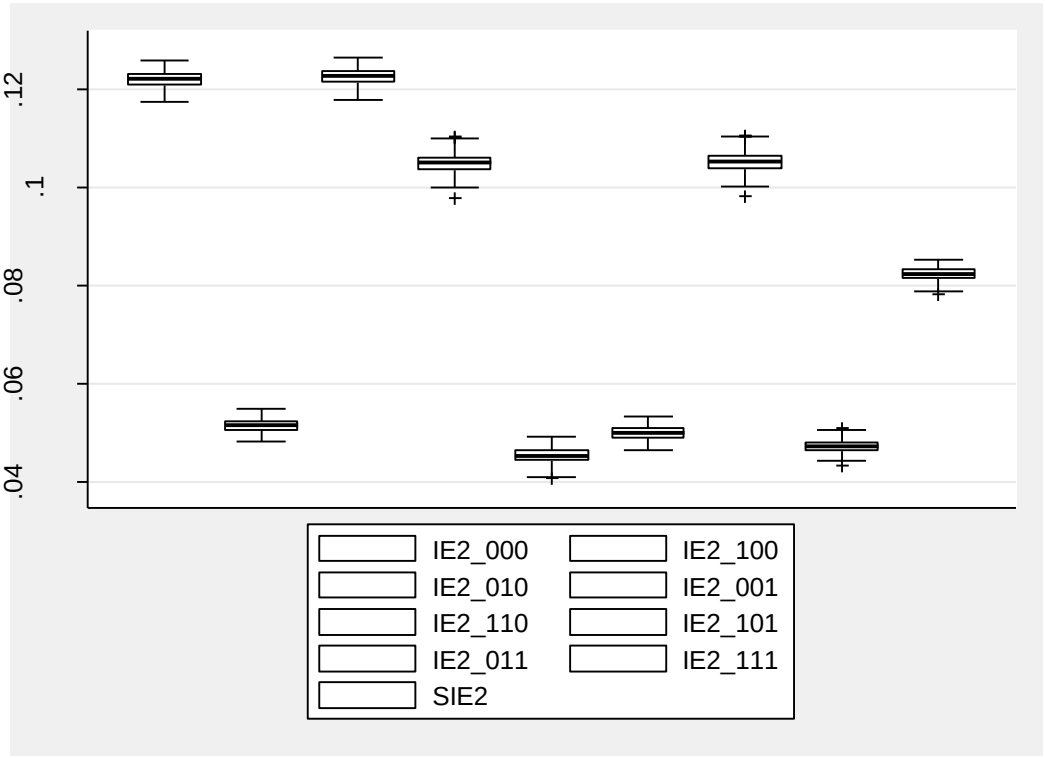


Figure 7/a NIE_2 Duty to obey – Study 2

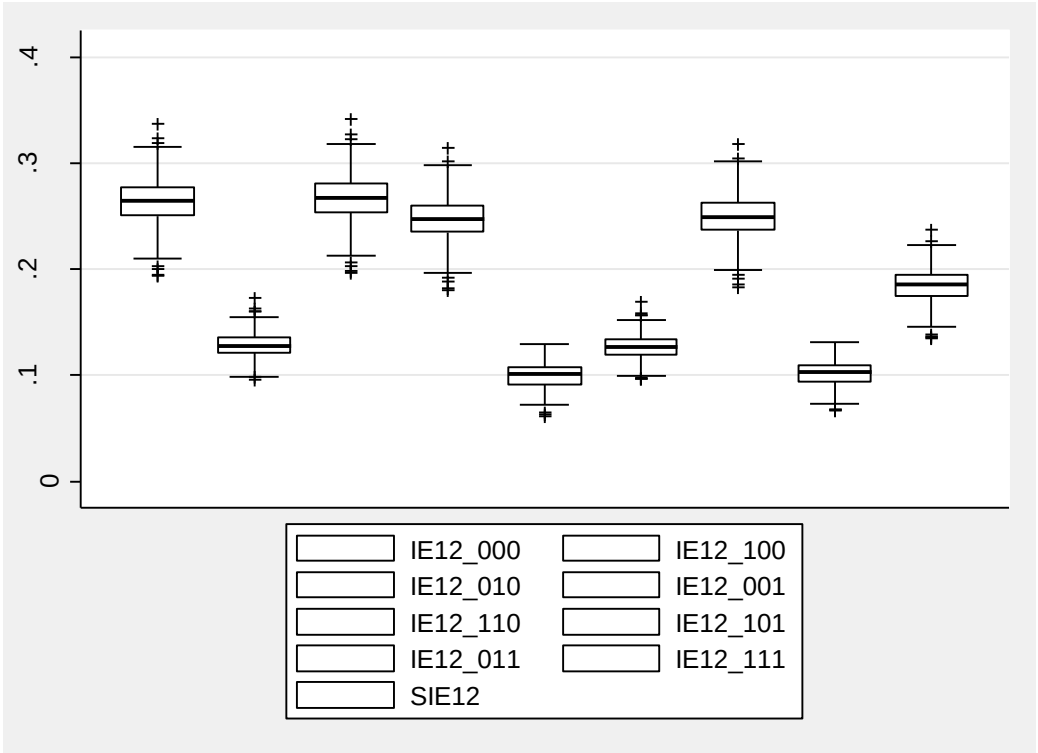


Figure 8/a NIE_{12} Joint effect – Study 2

References

Antoun, Christopher, Chan Zhang, Frederick G. Conrad, and Michael F. Schober. 2016. "Comparisons of Online Recruitment Strategies for Convenience Samples : AdWords , Facebook , and Amazon Mechanical Turk." *Field Methods* 28(3):231–46.

Avin, Chen, Ilya Shpitser, and Judea Pearl. 2005. "Identifiability of Path-Specific Effects." *Proceedings of the International Joint Conferences on Artifical Intelligence* 34:163–92.

Baron, Reuben M. and David A. Kenny. 1986. "Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6):173–82.

Bentler, P. M. and C. P. Chou. 1987. "Practical Issues in Structural Modeling." *Sociological Methods and Research* 16(1):78–117.

Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2016. "Can We Turn Shirkers into Workers?" *Journal of Experimental Social Psychology* 66:20–28.

Bollen, Kenneth A. and Judea Pearl. 2013. "Eight Myths About Causality and Structural Equation Models." Pp. 301–28 in *Handbook of Causal Analysis for Social Research*, edited by S. L. Morgan. Springer.

Bradford, Ben. 2014. "Policing and Social Identity: Procedural Justice, Inclusion and Cooperation between Police and Public." *Policing and Society* 24(1):22–43.

Bradford, Ben, Elizabeth Stanko, and Jonathan Jackson. 2012. *Just Authority? - Trust in the Police in England and Wales*. Routledge.

Browne, Michael W. and Robert Cudeck. 1992. "Alternative Ways of Assessing Model Fit." *Sociological Methods & Research* 21(2):230–58.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk." *Perspectives on Psychological Science* 6(1):3–5.

Chandler, Jesse, Pam Mueller, and Gabriele Paolacci. 2014. "Nonnaïveté among Amazon Mechanical Turk Workers: Consequences and Solutions for Behavioral Researchers." *Behavior Research Methods* 46(1):112–30.

Coffman, D. L. and W. Zhong. 2012. "Assessing Mediation Using Marginal Structural Models in the Presence of Confounding and Moderation." *Psychological Methods* 17(4):642–64.

Daniel, R. M., B. L. De Stavola, S. N. Cousens, and S. Vansteelandt. 2015. “Causal Mediation Analysis with Multiple Mediators.” *Biometrics* 71(1):1–14.

Daniel, Rhian M., Bianca L. De Stavola, and Simon N. Cousens. 2011. “Gformula - Estimating Causal Effects in the Presence of Time-Varying Confounding or Mediation Using the g-Computation Formula.” *The Stata Journal* 11(4):479–517.

Daniel, Rhian M., Bianca L. De Stavola, and Stijn Vansteelandt. 2016. “The Formal Approach to Quantitative Causal Inference in Epidemiology: Misguided or Misrepresented?” *International Journal of Epidemiology* 45(6):1817–29. Retrieved (http://fdslive.oup.com/www.oup.com/pdf/production_in_progress.pdf).

Greenland, Sander. 2017. “For and Against Methodologies: Some Perspectives on Recent Causal and Statistical Inference Debates.” *European Journal of Epidemiology* 32(1):1–18.

Hauser, David J. and Norbert Schwarz. 2016. “Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants.” *Behavior Research Methods* 48(1):400–407.

Horton, John J., David G. Rand, and Richard J. Zeckhauser. 2011. “The Online Laboratory: Conducting Experiments in a Real Labor Market.” *Experimental Economics* 14(3):399–425.

Hough, Mike, Jonathan Jackson, and Ben Bradford. 2013. “Legitimacy, Trust and Compliance: An Empirical Test of Procedural Justice Theory Using the European Social Survey.” Pp. 326–53 in *Legitimacy and Criminal Justice - An International Exploration*, edited by J. Tankebe and A. Liebling. Oxford University Press.

Huq, A. Z. Aziz H., J. Jackson, and R. J. Trinker. 2017. “Legitimating Practices: Revisiting the Predicates of Police Legitimacy.” *British Journal of Criminology* (57):1101–22.

Imai, K., D. Tingley, and T. Yamamoto. 2013. “Experimental Designs for Identifying Causal Mechanisms.” *Journal of the Royal Statistical Society Series A-Statistics in Society* 176(1):5–51.

Imai, Kosuke and Luke Keele. 2014. “‘Comment on Pearl: Practical Implications of Theoretical Results for Causal Mediation Analysis’: Correction to Imai et Al. (2014).” *Psychological Methods* 19(4):482–87.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. “A General Approach to Causal Mediation Analysis.” *Psychological Methods* 15(4):309–34.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. “Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies.” *American Political Science Review* 105(4):765–89.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010a. “Identification, Inference and Sensitivity Analysis for Causal Mediation Effects.” *Statistical Science* 25(1):51–71. Retrieved (http://arxiv.org/abs/1011.1079).

Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010b. “Identification, Inference and Sensitivity Analysis for Causal Mediation Effects.” *Statistical Science* 25(1):51–71.

Imai, Kosuke and Teppei Yamamoto. 2013. “Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments.” *Political Analysis* 21(2):141–71.

Jackson, Jonathan et al. 2012. “Why Do People Comply with the Law?” *British Journal of Criminology* 52(6):1051–71.

Jo, Booil. 2008. “Causal Inference in Randomized Experiments With Mediational Processes.” *Psychological Methods* 13(4):314–36.

Judd, Charles M. and David A. Kenny. 1981. *Estimating the Effects of Social Interventions*. Camdridge University Press.

Kang, Joseph, Xiaogang Su, Lei Liu, and Martha L. Daviglus. 2014. “Causal Inference of Interaction Effects with Inverse Propensity Weighting, G-Computation and Tree-Baes Standardization.” *Statistical Analysis and Data Mining* 7:323–36.

Kaplan, David. 2008. *Structural Equation Modeling - Foundations and Extensions*. 2nd ed. SAGE.

Keele, Luke. 2015a. “Causal Mediation Analysis Warning! Assumptions Ahead.” *American Journal of Evaluation* 46(4):500–513.

Keele, Luke. 2015b. “The Statistics of Causal Inference: A View from Political Methodology.” *Political Analysis* 23(3):313–35.

Kennedy, Edward H. 2015. “Semiparametric Theory and Empirical Processes in Causal Inference.” 1–26. Retrieved (http://arxiv.org/abs/1510.04740).

Kenny, David A. 2008. “Reflections on Mediation.” *Organizational Research*

Methods 11(2):353–58.

Kim, Chanmin, Michael J. Daniels, and Joseph W. Hogan. 2018. "Bayesian Methods for Multiple Mediators : Relating Principal Stratification and Causal Mediation in the Analysis of Power Plant Emission Controls." *Biostatistics* In press:1–36.

Kline, Rex B. 2015. "The Mediation Myth." *Basic and Applied Social Psychology* 37(4):202–13.

Kuha, Jouni and John H. Goldthorpe. 2010. "Path Analysis for Discrete Variables: The Role of Education in Social Mobility." *Journal of the Royal Statistical Society Series A-Statistics in Society* 173:351–69.

Lange, Theis, Mette Rasmussen, and Lau Caspar Thygesen. 2014. "Assessing Natural Direct and Indirect Effects through Multiple Pathways." *American Journal of Epidemiology* 179(4):513–18.

Liu, Pengfei, Ji Chen, Zhaohua Lu, and Xinyuan Song. 2015. "Transformation Structural Equation Models With Highly Nonnormal and Incomplete Data." *Structural Equation Modeling: A Multidisciplinary Journal* 22(3):401–15.

Loeys, Tom, Beatrijs Moerkerke, An Raes, Yves Rosseel, and Stijn Vansteelandt. 2014. "Estimation of Controlled Direct Effects in the Presence of Exposure-Induced Confounding and Latent Variables." *Structural Equation Modeling* 21(3):396–407.

Mackinnon, David P. 2008. *Introduction to Statistical Mediation*. Erlbaum.

Mackinnon, David P., Yasemin Kisbu-sakarya, and Amanda C. Gottschall. 2013. "Developments in Mediation Analysis Oxford Handbooks Online Developments in Mediation Analysis." Pp. 1–28 in *Oxford Handbook of Quantitative Methods*, vol. 2, edited by T. D. Little. New York: Oxford University Press.

Mackinnon, David P. and Angela G. Pirlott. 2015. "Statistical Approaches for Enhancing Causal Interpretation of the M to Y Relation in Mediation Analysis." *Personality and Social Psychology Review* 19(1):30–43.

Maglio, Sam J., Yaacov Trope, and Nira Liberman. 2013. "The Common Currency of Psychological Distance." *Current Directions in Psychological Science* 22(4):278–82.

Malhotra, Neil. 2008. "Completion Time and Response Order Effects in Web Surveys." *Public Opinion Quarterly* 72(5):914–34.

Manski, Charles F. 2007. *Identification for Prediction and Decision*. Harvard

University Press.

Mason, Winter and Siddharth Suri. 2012. “Conducting Behavioral Research on Amazon’s Mechanical Turk.” *Behavior Research Methods* 44(1):1–23.

Mauro, Robert. 1990. “Understanding L.O.V.E. (Left Out Variable Error): A Method for Estimating the Effects of Omitted Variables.” 108(2):314–29.

Mayer, Axel, Nora Umbach, Barbara Flunger, and Augustin Kelava. 2017. “Effect Analysis Using Nonlinear Structural Equation Mixture Modeling.” *Structural Equation Modeling: A Multidisciplinary Journal* 24(4):556–70.

Mazerolle, Lorraine, Emma Antrobus, Sarah Bennett, and Tom R. Tyler. 2013. “Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice.” *Criminology* 51(1):33–63.

Moerkerke, Beatrijs, Tom Loeys, and Stijn Vansteelandt. 2015. “Structural Equation Modeling versus Marginal Structural Modeling for Assessing Mediation in the Presence of Posttreatment Confounding.” *Psychological Methods* 20(2):204–20.

Mutz, Diana C. and Robin Pemantle. 2016. “Standards for Experimental Research : Encouraging a Better Understanding of Experimental Methods.” *Journal of Experimental Political Science* 192–215.

Oppenheimer, Daniel M., Tom Meyvis, and Nicolas Davidenko. 2009. “Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power.” *Journal of Experimental Social Psychology* 45(4):867–72.

Paolacci, Gabriele, Jesse Chandler, and Pg Ipeirotis. 2010. “Running Experiments on Amazon Mechanical Turk.” *Judgment and Decision Making* 5(5):411–19.

Pearl, Judea. 2010. “On the Consistency Rule in Causal Inference: Axiom, Definition, Assumption, or Theorem?” *Epidemiology* 21(6):872–75.

Pearl, Judea. 2014. “Interpretation and Identification of Causal Mediation.” *Psychological Methods* 19(4):459–81.

Peer, Eyal, Joachim Vosgerau, and Alessandro Acquisti. 2013. “Reputation as a Sufficient Condition for Data Quality on Amazon Mechanical Turk.” *Behavior Research Methods* 1023–31.

Petersen, Maya L., Sandra E. Sinisi, and Mark J. van der Laan. 2006. “Estimation of Direct Causal Effects.” *Epidemiology* 17(3):276–84.

Pirlott, Angela G. and David P. Mackinnon. 2016. “Design Approaches to Experimental Mediation ☆.” *Journal of Experimental Social Psychology* 66:29–38.

Preacher, Kristopher J. 2015. “Advances in Mediation Analysis: A Survey and Synthesis of New Developments.” *Annual Review of Psychology* 66:825–52.

Rivera, Lauren A. and Andras Tilcsik. 2016. “Class Advantage, Commitment Penalty: The Gendered Effect of Social Class Signals in an Elite Labor Market.” *American Sociological Review* 1–68.

Robins, James M. 1987. “A Graphical Approach to the Identification and Estimation of Causal Parameters in Mortality Studies with Sustained Exposure Periods.” *Journal of Chronic Diseases* 40(2):139–61.

Robins, James M. and Sander Greenland. 1992. “Identifiability and Exchangeability for Direct and Indirect Effects.” *Epidemiology* 3(2):143–55.

Sardeshmukh, Shruti R. and Robert J. Vandenberg. 2016. “Integrating Moderation and Mediation : A Structural Equation Modeling Approach.” 1–25.

Shadish, William R., Thomas D. Cook, and Donald Thomas Campbell. 2002. *Experimental and Quasi-Experimental Design for Generalized Causal Inference*. Houghton Mifflin Company.

Snowden, Jonathan M., Sherri Rose, and Kathleen M. Mortimer. 2011. “Implementation of G-Computation on a Simulated Data Set: Demonstration of a Causal Inference Technique.” *American Journal of Epidemiology* 173(7):731–38.

De Stavola, Bianca L., Rhian M. Daniel, George B. Ploubidis, and Nadia Micali. 2015. “Mediation Analysis with Intermediate Confounding: Structural Equation Modeling Viewed through the Causal Inference Lens.” *American Journal of Epidemiology* 181(1):64–80.

Steen, Johan, Tom Loeys, Beatrijs Moerkerke, and Johan Steen. 2017. “Flexible Mediation Analysis with Multiple Mediators.” *American Journal of Epidemiology* 186(2):184–93.

Steen, Johan, Tom Loeys, Beatrijs Moerkerke, and Stijn Vansteelandt. 2017. “Medflex : An R Package for Flexible Mediation Analysis Using Natural Effect Models.” *Journal of Statistical Software* 76(11):1–45.

Steiner, Peter M., Thomas D. Cook, William R. Shadish, and M. H. Clark. 2010. “The Importance of Covariate Selection in Controlling for Selection Bias in Observational Studies.” 15(3):250–67.

Taguri, Masataka, John Featherstone, and Jing Cheng. 2017. “Causal Mediation Analysis with Multiple Causalla Non-Ordered Mediators.” *Statistical Methods*

in Medical Research 1–21.

Tchetgen Tchetgen, Eric J. and Tyler J. VanderWeele. 2014. "Identification of Natural Direct Effects When a Confounder of the Mediator Is Directly Affected by Exposure." *Epidemiology.* 25(2):282–91.

Tomarken, Andrew J. and Niels G. Waller. 2005. "Structural Equation Modeling: Strengths, Limitations, And Misconceptions." *Annual Review of Clinical Psychology* 1:31–65.

Tourangeau, Roger, Mick P. Couper, and Frederick G. Conrad. 2013. "Up Means Good: The Effect of Screen Position on Evaluative Ratings in Web Surveys." *Public Opinion Quarterly* 77(S1):69–88.

Trinkner, Rick, Jonathan Jackson, and Tom R. Tyler. 2017. "Bounded Authority: Expanding 'Appropriate' Police Behavior Beyond Procedural Justice." *Law and Human Beh* 42(3):280–93.

Tyler, Phillip Atiba Goff, and Robert J. MacCoun. 2015. "The Impact of Psychological Science on Policing in the United States: Procedural Justice, Legitimacy, and Effective Law Enforcement." *Psychological Science in the Public Interest* 16(3):75–109.

Tyler, Tom and Jeffrey Fagan. 2008. "Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities?" *Ohio State Journal of Criminal Law* 6:231–75.

Tyler, Tom R. 2006. *Why People Obey the Law*. Princeton: Princeton University Press.

Tyler, Tom R. and Jonathan Jackson. 2013. "Future Challenges in the Study of Legitimacy and Criminal Justice." Pp. 83–104 in *Legitimacy and Criminal Justice - An International Exploration,* edited by J. Tankebe and A. Liebling. Wiley.

Tyler, Tom R. and Jonathan Jackson. 2014. "Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation, and Engagement." *Psychology, Public Policy, and Law* 20(1):78–95.

VanderWeele, Tyler J. 2012. "Invited Commentary: Structural Equation Models and Epidemiologic Analysis." *American Journal of Epidemiology* 176(7):608–12.

VanderWeele, Tyler J. 2015. *Explanation in Causal Inference - Methods for Mediation and Interaction*. Oxford University Press.

VanderWeele, Tyler J. 2016. "Mediation Analysis: A Practitioner's Guide." *Annual*

Review of Public Health 37(1):17–32.

VanderWeele, Tyler J. and Mirjam J. Knol. 2014. "A Tutorial on Interaction." *Epidemiological Methods* 3(1):33–72.

VanderWeele, Tyler J. and Stijn Vansteelandt. 2014. "Mediation Analysis with Multiple Mediators." *Epidemiologic Methods* 2(1):95–115.

Vansteelandt, Stijn and Rhian M. Daniel. 2017. "Interventional Effects for Mediation Analysis with Multiple Mediators." *Epidemiology* 28(2):258–65.

Vansteelandt, Stijn and Niels Keiding. 2011. "Invited Commentary: G-Computation-Lost in Translation?" *American Journal of Epidemiology* 173(7):739–42.

Wang, Aolin and Onyebuchi A. Arah. 2015a. "G-Computation Demonstration in Causal Mediation Analysis." *European Journal of Epidemiology* 30(10):1119–27.

Wang, Aolin and Onyebuchi A. Arah. 2015b. "G-Computation Demonstration in Causal Mediation Analysis." *European Journal of Epidemiology* 30(10):1119–27.

Wang, Aolin, Roch A. Nianogo, and Onyebuchi A. Arah. 2017. "G-Computation of Average Treatment Effects on the Treated and the Untreated." *BMC Medical Research Methodology* 17(1):1–5.

Westreich, Daniel et al. 2015. "Imputation Approaches for Potential Outcomes in Causal Inference." *International Journal of Epidemiology* (July):1731–37.

White, Michael D., Philip Mulvey, and Lisa M. Dario. 2016. "Arrestees' Perceptions of the Police." *Criminal Justice and Behavior* 43(3):343–64.

Concluding Remarks

In this concluding chapter, I synthesise the main findings from the four papers of the thesis. I incorporate them into a more cohesive model that can be compared to the comprehensive model outlined in the introductory chapter. I discuss the limitations of the present thesis and suggest future directions of research.

Summary of the findings

This thesis has provided an examination of the comprehensive model set out in the introductory chapter. Each paper focussed on the treatment effects of either contact with the police (Paper 1 and Paper 3) or procedural justice and respect for boundaries of the police (Paper 2 and Paper 4). In each paper, the mediators varied. Paper 1 only examined procedural justice. Paper 2 scrutinised mediators with post-treatment confounding and their indirect effects on police and legal legitimacy. In Study 1, sense of power, police grip on power, and social identification were examined, which were complemented by self-control in Study 2, with Study 3 solely testing sense of power as the mediator using an alternative design-based estimation. Paper 3 took two sequential mediators, with procedural justice and sense of power as first, and free duty to obey the police and coercive obligation to obey the police as second mediators of the impact of contact on compliance with the law and willingness to cooperate with the police. Notably, in the appendix for Paper 3, another sequential model was fitted, which resembled the comprehensive model more closely, placing procedural justice as the first, sense of power as the second, and a version of duty to obey as the third mediator. Finally, Paper 4 took normative alignment with and duty to obey the police as mediators which either confounded one-another or were assumed to follow a causal order, with normative alignment affecting duty to obey.

The results from these four papers are summarised by Figure 1, which only includes the pathways that were reasonably strong and/or present in multiple papers. This thesis has at least four notable findings. First, previous experiences with the police (measured as contact with the police) had an impact on procedural justice, which in turn mediated its impact towards duty to obey the police and normative alignment with the police (Paper 1). Moreover, the impact of previous experiences was mediated on legal compliance and willingness to cooperate with the police by multiple mediators

(i.e., procedural justice, sense of power, free/coercive obligation, Paper 3). This implies that (1) in line with a plethora of literature (e.g., Bradford 2017; Gau 2013; Lowrey, Maguire, and Bennett 2016; Tyler et al. 2014), the quality of contact affects procedural justice and (2) this contact has crucial downstream effects with procedural justice, sense of power, and legitimacy as potential intermediate variables.

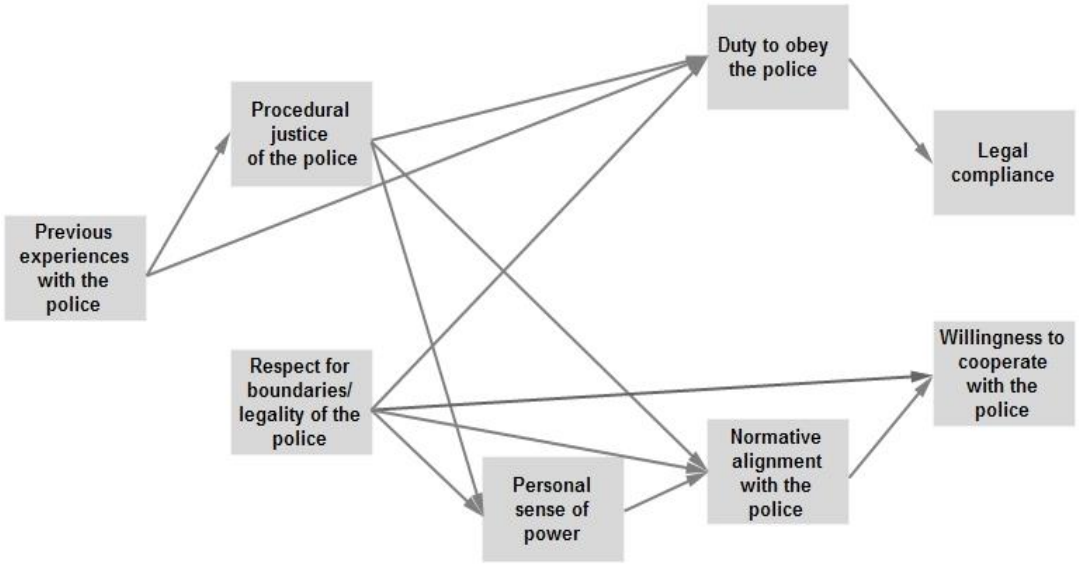


Figure 1 Summary of this thesis's main findings

Second, the findings seem to support the central role of procedural justice in the theory. Procedural justice not only mediated the effect of previous contact on both aspects of legitimacy (Paper 1), or had a direct impact on them (Paper 4), but for the most part, it continued to influence them even after considering the psychological mechanisms in play (Paper 2). Importantly, however, in the papers where societally desirable outcomes (i.e., legal compliance and willingness to cooperate) were included, the impact of procedural justice dropped significantly or was not different from zero when one or the other aspects of legitimacy was introduced as a mediator (Paper 2, Paper 3, and Paper 4). This implies that the impact of procedural justice on cooperation and compliance is by and large absorbed by police legitimacy, which in turn transmits its effects. Thus, the causal evidence suggests that expected fair and respectful treatment by the police which allows voice and communicates trustworthy motives is crucial in shaping (1) people's judgments regarding the moral appropriateness of police behaviour (normative alignment), (2) citizens' evaluation of

whether they should give consent to police actions even when they disagree with them (duty to obey), and that (3) the impact of procedural justice on legal compliance and cooperation with the police is to a large extent/fully mediated by evaluations regarding the legitimacy of the police.

Conspicuously, procedural justice and respect for boundaries are separated in the Figure 1, because the latter was only measured in Paper 2 and Paper 4. However, when respect for boundaries was included, both procedural justice and respect for boundaries were highly correlated and moved around in a similar fashion as a result of the experimental manipulation. Hence, arguably, the findings of Paper 2 and Paper 4 also apply to respect for boundaries, thus those pathways are attributed to both procedural justice and respect for boundaries. Future empirical and theoretical work should determine whether these two attributes of appropriate police behaviour should be kept separate or integrated as one construct.

Third, and as a relative novelty, sense of power was identified as a mediator for the impact of procedural justice on both normative alignment with the police and the law (Paper 2). Moreover, the impact of previous contact on willingness to cooperate and compliance with the law was also mediated by sense of power in the absence of legitimacy (Paper 3). Even after considering various other psychological processes, sense of power remained the only construct that mediated the impact of procedural justice and respect for boundaries on legitimacy of the law and the police (Paper 2). The causal evidence seems to support the claim that positive contact, the perceived procedurally fair treatment and respect for boundaries by the police, increases an individual's sense of control and power in expected future interactions with the police. This increased sense of power in turn channels (1) the impact of the previous experience (i.e., contact) on societally desirable outcomes (i.e., cooperation and compliance) when legitimacy is not present and (2) the impact of procedural justice on people's evaluation of whether the police and the law cherish the same rights and wrong as they do (i.e., normative alignment with the police and the law). Personal sense of power mediates the effect of procedural justice either due to instrumental considerations (i.e., people legitimate authority figures they feel they have power over) or normative/relational assessments (i.e., people legitimate authority figures that they associate with consent not coercion), or both. Future studies should attempt to replicate these findings and disentangle these two aspects.

It is worth mentioning the absence of social identification from Figure 1. Procedural justice only had a tentative indirect effect on social identification, whilst the previous contact had a much larger direct effect which was not channelled by procedural justice (Paper 1). This tentative indirect effect was called into question in two experiments (Study 1 and Study 2, Paper 2), where manipulating procedural justice and respect for boundaries produced no or weak average treatment effects on social identification, which in turn meant that social identification did not mediate the effect of procedural justice on either of the legitimacy variables. A closer examination of alternative explanations for the lack of findings indicated that social identification did have a significant partial association with police and legal legitimacy, but those effects did not arise from or vary by the treatment and as such no causal properties could be attributed to them (Appendix/D, Paper 2). The lack of causal relationship makes it plausible that social identification is merely a correlate of police legitimacy and that it is not informed by procedural justice. Admittedly, this startling finding contradicts much of the experimental work done in organisational settings (Blader and Chen 2012; Blader and Tyler 2009; Tyler 2017) and requires further support from future studies. It might also be worth considering alternative conceptualisations and operationalisations of social identification, such as is found in the literature on preference formation in political science (Kalin and Sambanis 2018) or the group entitativity and density models in social psychology (Alves, Koch, and Unkelbach 2016; Greenaway et al. 2016).

Finally, the two aspects of police legitimacy appeared to influence distinct societally desirable outcomes. Even though both duty to obey the police and normative alignment with the police were influenced by previous contact and procedural justice (Paper 1-Paper 4), duty to obey appeared to predict legal compliance (Paper 3), whilst normative alignment seemed to have an effect on willingness to cooperate (Paper 4). The evidence from Paper 3 and Paper 4 might seem contradictory as Paper 3 implied that duty to obey mediated the impact of contact on cooperation. However, the more rigorous test of Paper 4, where procedural justice and legality were manipulated and both aspects of legitimacy were included, did not find this mediated effect on cooperation. These causal results also appear to support earlier observational findings and theoretical work based on which normative alignment was identified as the proactive aspect of legitimacy associated with increased community engagement and cooperation, whilst duty to obey was considered the reactive aspect associated with

legal compliance and acceptance of rightful police use of force (e.g., Bradford, Milani, and Jackson 2017; Jackson 2018; Moravcová 2016; Tyler and Jackson 2014). All in all, from the two aspects of police legitimacy, the results suggest that (1) the perceived moral appropriateness of the police (normative alignment) causally mediates the impact of perceived fairness and respect for boundaries towards willingness to cooperate with the police, (2) the willing consent to police actions (duty to obey) causally mediates the influence of contact on future legal compliance, and (3) despite some suggestions in the literature (Hamm, Trinkner, and Carr 2017; Huq, Jackson, and Trinker 2017), there is no causal evidence for the sequential order of the two aspects when the impact of procedural justice is transmitted towards willingness to cooperate. The two aspects of legal legitimacy were not included in Figure 1 as they were only measured as outcomes in Paper 2.

These four main findings provide answers to all the research questions set out in the literature review:

Q1 Procedural justice does appear to mediate the impact of previous contact, although its effects are stronger and more robust on police legitimacy than on social identification (Paper 1).

Q2 From the psychological processes surveyed in Paper 2, only personal sense of power emerged as a significant mediator of the influence of procedural justice and respect for boundaries, showing relatively robust and enduring indirect effects across the three studies.

Q3 From the three mediators, free duty to obey mediated the effect of contact on cooperation with the police and compliance with the law, whilst procedural justice and sense of power mediated the effects in the absence of free duty to obey the police (Paper 3).

Q4 Finally, from the two aspects of police legitimacy, only normative alignment with the police mediated the impact of procedural justice and respect for boundaries on willingness to cooperate, regardless of the method of estimation pursued (Paper 4).

Lastly, it is worth taking stock of the different causal mediation analysis techniques used in the four papers. Paper 1 used causal mediation analysis with a single mediator, which was proven to be a versatile tool that allows more flexible modelling

than the usual alternatives in the SEM literature. Moreover, the clearly spelt out identifying assumption and the sensitivity analysis techniques can also help in assessing whether such analysis is feasible and regarding the robustness of the results. However, the major limitation of this approach is that it can only be applied to relatively simplistic models where a single mediator is considered. Such models are uncommon in the social sciences, thus this approach has limited value for most applications.

Yet, expanding the analysis to multiple mediators provides its unique challenges. Paper 4 reviewed the four alternative options one can pursue:

1. Assume causal independence of the various mediators, which permits estimating the pathways one at a time. This is usually untenable in the social sciences where mediators tend to be related to each other (i.e., there are intertwined pathways).
2. Resort to the estimation of joint effects of various mediators thus collating their impact (Paper 3). This is an easy solution, but it does not permit the decomposition of the various effects, which limits the information that could be gained by it.
3. Consider the mediators as post-treatment confounders of each other and estimate their effects one at a time (Paper 2 and Paper 4). From the two solutions presented in this thesis, the semi-parametric model used by Paper 2 is easier to estimate, but it relies on assumptions that are more difficult to satisfy. By contrast, the g-computation approach taken by Paper 4 is more flexible but more difficult to estimate.
4. Impose a sequential order where some sort of decomposition becomes feasible (Paper 3 and Paper 4). In case of the natural effects model of Paper 3, the sequential approach requires weaker assumptions and permits more flexible modelling, but it also attributes the jointly mediated effects to the first mediator. By contrast, the g-computation solution has stronger assumptions and more difficult estimation but provides the finest possible decomposition (separate and jointly mediated effects).

Ultimately, and as with any other methodological consideration, the solution chosen by a researcher should be guided by theory and a research question. For instance, in a

policy evaluation, one might not be able to (or want to) satisfy all the assumptions required by options 3 and 4 and estimate the two aspects of legitimacy separately. This researcher could still rely on option 2 as a straightforward solution, and assess police legitimacy and its mediating role as a whole.

Study 3 in Paper 2 was unique as it used a design-based alternative to estimate the causally mediated effects. As discussed there, it can be very difficult to figure out which assumptions and corresponding estimation methods to rely upon. Moreover, the results gained were also more uncertain with very wide confidence intervals and smaller effect sizes than reported in the other two studies in Paper 3. Unfortunately, I am not aware of simulation studies on the statistical power calculation for parallel and parallel encouragement designs, but it is conceivable that an even larger sample would have been needed to further reduce the standard errors. With all these considered and despite all the difficulties that the design-based approach entails, I hope that this paper inspires other researchers to use similar innovative research designs to assess causal mechanisms in the future.

Limitations

I should nevertheless address some limitations of this thesis. First, there are questions regarding the generalisability of the results. Both Paper 1 and Paper 3 relied on the dataset collected during the Scottish Community Engagement Trial (ScotCET). Paper 1 put a great deal of effort into re-assessing the apparent implementation failure and found that the treatment effects are attributable to the research design. With this considered, the sample was still very peculiar. The ScotCET data utilised mostly mundane traffic encounters in a very high-trust community in rural Scotland with less than 7% of the questionnaires returned (MacQueen and Bradford 2015). This means that the results from Paper 1 and Paper 3 probably only apply to a subpopulation of similar traffic encounters.

By contrast, Paper 2 and Paper 4 recruited participants from either Amazon Turk or Prolific Academic. Several studies have shown that online crowdsourced samples are relatively diverse (Antoun et al. 2016; Paolacci and Chandler 2014; Ross et al. 2010), that much research can be replicated online (Coppock 2018; Horton, Rand, and Zeckhauser 2011), and that they tend to provide good data quality (Hauser and Schwarz 2016; Lovett et al. 2018; Peer et al. 2017). However, these are still large convenience samples where people self-select to certain studies, which means that the

external validity of the findings is limited. As a further issue, the experimenter has relatively little control over the subjects; this emerged as a clear problem in the parallel (encouragement) design in Paper 2. Accordingly, it is of primary importance that the experimental procedures from these four papers are replicated relying on other randomised controlled trials drawn from other populations or on experiments that were carried out in a controlled environment such as a laboratory, but also to find similar findings on the same or similar populations to confirm that the results did not emerge due to a mere statistical fluke.

Second, one might criticise the relatively wide scope of this thesis. Instead of focussing on a single part of the comprehensive model outlined in the introduction (e.g., only on the psychological mediators of the impact of procedural justice on various outcomes), this thesis undeniably casts a relatively wide net and touched upon all of the components of that model. The motivation for this was twofold. First, the wider literature routinely tests models of similar complexity using structural equation modelling (e.g., Huq et al. 2017; Jackson et al. 2012; Murphy, Bradford, and Jackson 2016). Second, the demonstration of the alternative techniques in the causal mediation analysis literature also required testing models with several conceptual layers. This effort is exemplified by Paper 3, which on its own touched upon all elements of the comprehensive model using a natural effects model with sequentially ordered mediators. Alas, this holistic approach also meant that instead of interrogating one topic, the thesis only assessed each component in a more cursory fashion.

Third, and as a further complication for the work presented here, the causal interpretation of the results is dependent on some untestable assumptions. Despite the potential influence of unmeasured pre-treatment confounding being quantifiable (as demonstrated in Paper 1, Paper 2, and Paper 4) and a wide range of covariates being included (especially in Paper 1-Paper 3), there is still a possibility that the relationships found here are only spurious and caused by a pre-treatment confounder that has not been accounted for. It is equally likely that an unmeasured post-treatment confounder might explain some of the results. For instance, it is possible that beyond personal sense of power an individual's emotions and emotion regulation could also transmit the impact of procedural justice towards legitimacy of the police (Johnson et al. 2016; Yesberg and Bradford 2018). Nevertheless, the sensitivity analyses techniques applied here can make the newly arising results from other studies comparable, which in the

future can provide a better understanding of the robustness of the findings of this thesis.

Finally, a further issue is that for the most part the causal order was imposed on the different models without the temporal order being established. As argued in the introduction, the conceptual framework was informed by the available evidence and the principles of cognitive information processing, presuming that quicker effortless processes are more fundamental than slower more elaborate ones (Barclay, Bashshur, and Fortin 2017; Von Hippel, Lakin, and Shakarchi 2005; Kruglanski 1996). With the exception of sense of power which was manipulated using the parallel design, there is still a possibility that some of the potential mechanisms could have been assigned to a wrong place. In the future, causal mediation analysis on longitudinal datasets could further clarify each concept's place in the theory (Walters 2017; Walters and Mandracchia 2017).

Future directions of research

The work carried out throughout my PhD has opened multiple future avenues of research. To conclude, I would like to address a few instances where theoretical advancement can be complemented by other methods of causal mediation analysis that have not been used in this thesis.

First, an obvious next step is to apply Bayesian alternatives of the frequentist methods discussed here. Some theoretical work has argued that Bayesian inference comes closer to capturing human information processing regarding procedural justice (Augustyn 2016). But even without this consideration, the large accumulated evidence warrants methods that are capable of including prior information. Yet there is a dearth of research in the procedural justice literature which uses Bayesian techniques (for a notable exception see: Lowrey, Maguire, and Bennett 2016). Bayesian methods would permit testing the probability of the hypothesis given the data instead of the null hypothesis. Moreover, Bayesian methods tend to outperform their frequentist peers in case of moderate sample sizes (Mcneish 2016). There are various techniques available for a single causal mediator (Daniels et al. 2012; Kim et al. 2016; Miočević et al. 2018), but recently a novel approach has also been proposed for multiple mediators (Kim, Daniels, and Hogan 2018). The more information we gather on the mediated effects in the procedural justice literature, the more pressing it becomes to utilise Bayesian alternatives.

Second, it is also plausible that certain mediated effects are dependent on third variables. This consideration harkens back to the assessment of treatment effect and design heterogeneity carried out in Paper 1. For instance, the stop and search literature has established that the number of previous encounters shapes people's views about whether the most recent one was procedurally fair (Tyler et al. 2014). Thus, it is possible that an otherwise procedurally just encounter's effect (treatment) that is transmitted through the evaluation of procedural justice (mediator) on police legitimacy (outcome) might be dependent on the number of earlier encounters (moderator). The causal effects of such moderated mediation models require fewer assumptions to be estimable than other causal mediation analysis models (Loeys et al. 2016) and can inform policy regarding conditional (heterogeneous) indirect effects (Braga et al. 2014; Na, Loughran, and Paternoster 2015).

Third, it remains an open question whether the effects of training, randomised controlled trials and experiments manipulating the procedurally just behaviour of the police or the perception of procedural justice are consistently estimable or plagued by treatment noncompliance. It is well-documented that changing police practices is usually an arduous undertaking (Hassell and Lovell 2015; MacQueen and Bradford 2017; Worden and McLean 2017). However, this is not always entirely due to the resistance by the officers but also hindered by the persistence of interpersonal dynamics in certain communities. Imagine an initiative wherein an area with high dissatisfaction with the police half of the officers receives procedural justice training whilst half are allowed to carry on with their usual behaviour. Even when the police officers try to follow the principles learned at the training, they might encounter citizens who expect a different kind of behaviour and react and evaluate the situation in a negative way without giving a chance to the officers. Even worse, they might aggravate the situation by being hostile to the officer who may "lose their cool" and act in a procedurally unjust manner (Hough 2012). Thus, in such situations, the individuals might inadvertently self-select into a procedurally unjust condition. This is a very clear example of treatment noncompliance, which might require the researcher to rely on local direct and indirect effects derived for the compliers (Keele, Tingley, and Yamamoto 2015; Yamamoto 2014) instead of the regular average effects for the whole population. Even if the researchers have a relatively high level of certainty that noncompliance is not an issue, testing for noncompliance can be carried out as a further robustness check (Rosenbaum 2010).

Finally, other design-based causal mediation analysis techniques could also be used. A prime candidate is the crossover (encouragement) design where the same participant consecutively receives both the treatment and control condition whilst holding the mediator at the same level (Imai, Tingley, and Yamamoto 2013). For instance, the same participant could view and answer questions regarding two subsequent videos of police officers treating someone in procedurally just and unjust ways (treatment), whilst for half of them, personal sense of power (mediator) is kept constant (low or high) for both experiments. As with the parallel (encouragement) design, this could inform research about the mediating role of sense of power. In addition, however, such effects become estimable not only between individuals but within the same individual providing a better control for the estimated effects.

References for the concluding remarks

Alves, Hans, Alex Koch, and Christian Unkelbach. 2016. "My Friends Are All Alike - the Relation between Liking and Perceived Similarity in Person Perception." *Journal of Experimental Social Psychology* 62:103–17.

Antoun, Christopher, Chan Zhang, Frederick G. Conrad, and Michael F. Schober. 2016. "Comparisons of Online Recruitment Strategies for Convenience Samples : AdWords , Facebook , and Amazon Mechanical Turk." *Field Methods* 28(3):231–46.

Augustyn, Megan Bears. 2016. "Updating Perceptions of (In)Justice." *Journal of Research in Crime and Delinquency* 53(2):255–86.

Barclay, Laurie J., Michael R. Bashshur, and Marion Fortin. 2017. "Motivated Cognition and Fairness: Insights, Integration, and Creating a Path Forward." *Journal of Applied Psychology* 102(6):867–89.

Blader, Steven L. and Ya-Ru Chen. 2012. "Differentiating the Effects of Status and Power: A Justice Perspective." *Journal of Personality and Social Psychology* 102(5):994–1014.

Blader, Steven L. and Tom R. Tyler. 2009. "Testing and Extending the Group Engagement Model: Linkages between Social Identity, Procedural Justice, Economic Outcomes, and Extrarole Behavior." *Journal of Applied Psychology* 94(2):445–64.

Bradford, Ben. 2017. *Stop and Search and Police Legitimacy*. Routledge.

Bradford, Ben, Jenna Milani, and Jonathan Jackson. 2017. "Identity, Legitimacy and 'Making Sense' of Police Use of Force." *Policing: An International Journal of Police Strategies & Management* 40(3):614–27.

Braga, Anthony a., Christopher Winship, Tom R. Tyler, Jeffrey Fagan, and Tracey L. Meares. 2014. "The Salience of Social Contextual Factors in Appraisals of Police Interactions with Citizens: A Randomized Factorial Experiment." *Journal of Quantitative Criminology* 30(4):599–627.

Coppock, Alexander. 2018. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* In press:1–16. Retrieved (https://www.cambridge.org/core/product/identifier/S2049847018000109/type/journal_article).

Daniels, Michael J., Jason A. Roy, Chanmin Kim, Joseph W. Hogan, and Michael G.

Perri. 2012. “Bayesian Inference for the Causal Effect of Mediation.” *Biometrics* 68(4):1028–36.

Gau, Jacinta M. 2013. “Consent Searches as a Threat to Procedural Justice and Police Legitimacy: An Analysis of Consent Requests During Traffic Stops.” *Criminal Justice Policy Review* 24(6):759–77.

Greenaway, Katharine H., Tegan Cruwys, S. Alexander Haslam, and Jolanda Jetten. 2016. “Social Identities Promote Well-Being Because They Satisfy Global Psychological Needs.” *European Journal of Social Psychology* 46(3):294–307.

Hamm, J. A., R. Trinkner, and J. D. Carr. 2017. “Fair Process, Trust, and Cooperation: Moving Toward an Integrated Framework of Police Legitimacy.” *Criminal Justice and Behavior* 44(9):1183–1212.

Hassell, Kimberly D. and Rickie D. Lovell. 2015. “Fidelity of Implementation: Important Considerations for Policing Scholars.” *Policing and Society* 25(5):504–20.

Hauser, David J. and Norbert Schwarz. 2016. “Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants.” *Behavior Research Methods* 48(1):400–407.

Von Hippel, William, Jessica L. Lakin, and Richard J. Shakarchi. 2005. “Individual Differences in Motivated Social Cognition - The Case of Self-Serving Information Processing.” *Personality and Social Psychology Bulletin* 31(10):1347–57.

Horton, John J., David G. Rand, and Richard J. Zeckhauser. 2011. “The Online Laboratory: Conducting Experiments in a Real Labor Market.” *Experimental Economics* 14(3):399–425.

Hough, Mike. 2012. “Procedural Justice and Professional Policing in Times of Austerity.” *Criminology & Criminal Justice* 13(2):181–97.

Huq, A. Z. Aziz H., J. Jackson, and R. J. Trinker. 2017. “Legitimating Practices: Revisiting the Predicates of Police Legitimacy.” *British Journal of Criminology* (57):1101–22.

Imai, K., D. Tingley, and T. Yamamoto. 2013. “Experimental Designs for Identifying Causal Mechanisms.” *Journal of the Royal Statistical Society Series A-Statistics in Society* 176(1):5–51.

Jackson, Jonathan et al. 2012. “Why Do People Comply with the Law?” *British Journal of Criminology* 52(6):1051–71.

Jackson, Jonathan. 2018. “Norms, Normativity, and the Legitimacy of Justice Institutions: International Perspectives.” *Annual Reveiw of Law and Social Sciences* 14 In pres.

Johnson, Cathryn, Karen A. Hegtvedt, Nikki Khanna, and Heather L. Scheuerman. 2016. “Legitimacy Processes and Emotional Responses to Injustice.” *Social Psychology Quarterly* 79(2):95–114.

Kalin, Michael and Nicholas Sambanis. 2018. “How to Think About Social Identity.” *Annual Review of Political Science* 21(1):239–60.

Keele, Luke, Dustin Tingley, and Teppei Yamamoto. 2015. “Identifying Mechanisms behind Policy Interventions via Causal Mediation Analysis.” *Journal of Policy Analysis and Management* 34(4):937–63.

Kim, Chanmin, Michael J. Daniels, and Joseph W. Hogan. 2018. “Bayesian Methods for Multiple Mediators : Relating Principal Stratification and Causal Mediation in the Analysis of Power Plant Emission Controls.” *Biostatistics* In press:1–36.

Kim, Chanmin, Michael J. Daniels, Bess H. Marcus, and Jason A. Roy. 2016. “A Framework for Bayesian Nonparametric Inference for Causal Effects of Mediation.” *Biometrics* 73(2):401–9.

Kruglanski, Arie W. 1996. “Motivated Social Cognition - Principles of the Interface.” Pp. 493–520 in *Social Psychology: Handbook of Basic Principles,* edited by E. Higgins and A. W. Kruglanski. Guilford Press.

Loeys, Tom, Wouter Talloen, Liesbet Goubert, Beatrijs Moerkerke, and Stijn Vansteelandt. 2016. “Assessing Moderated Mediation in Linear Models Requires Fewer Confounding Assumptions than Assessing Mediation.” *British Journal of Mathematical & Statistical Psychology* 69(3):352–74.

Lovett, Matt, Saleh Bajaba, Myra Lovett, and Marcia J. Simmering. 2018. “Data Quality from Crowdsourced Surveys: A Mixed Method Inquiry into Perceptions of Amazon’s Mechanical Turk Masters.” *Applied Psychology* 67(2):339–66.

Lowrey, Belén V., Maguire, Edward R., and Bennett, Richard R. 2016. "Testing the Effects of Procedural Justice and Overaccommodation in Traffic Stops: A Randomized Experiment." *Criminal Justice and Behavior,* 43(10): 1430–1449.

MacQueen, Sarah and Ben Bradford. 2015. “Enhancing Public Trust and Police Legitimacy during Road Traffic Encounters: Results from a Randomised Controlled Trial in Scotland.” *Journal of Experimental Criminology* 11(3):419–43.

MacQueen, Sarah and Ben Bradford. 2017. “Where Did It All Go Wrong? Implementation Failure—and More—in a Field Experiment of Procedural Justice Policing.” *Journal of Experimental Criminology* 13(3):321–45.

Mcneish, Daniel. 2016. “On Using Bayesian Methods to Address Small Sample Problems On Using Bayesian Methods to Address Small Sample Problems.” *Structural Equation Modeling: A Multidisciplinary Journal* 23(5):750–73.

Miočević, Milica, Oscar Gonzalez, Matthew J. Valente, and David P. MacKinnon. 2018. “A Tutorial in Bayesian Potential Outcomes Mediation Analysis.” *Structural Equation Modeling* 25(1):121–36.

Moravcová, Eva. 2016. “Willingness to Cooperate with the Police in Four Central European Countries.” *European Journal on Criminal Policy and Research* 22(1):171–87.

Murphy, K., B. Bradford, and J. Jackson. 2016. “Motivating Compliance Behavior Among Offenders: Procedural Justice or Deterrence?” *Criminal Justice and Behavior* 43(1):102–18.

Na, Chongmin, Thomas A. Loughran, and Raymond Paternoster. 2015. “On the Importance of Treatment Effect Heterogeneity in Experimentally-Evaluated Criminal Justice Interventions.” *Journal of Quantitative Criminology* 31(2):289–310.

Nagin, Daniel S. and Cody W. Telep. 2017. “Procedural Justice and Legal Compliance.” *Annual Review of Law and Social Science* 13(1):5–28.

Paolacci, Gabriele and Jesse Chandler. 2014. “Inside the Turk: Understanding Mechanical Turk as a Participant Pool.” *Current Directions in Psychological Science* 23(3):184–88.

Peer, Eyal, Sonam Samat, Laura Brandimarte, and Alessandro Acquisti. 2017. “Beyond the Turk : Alternative Platforms for Crowdsourcing Behavioral Research.” *Journal of Experimental Social Psychology* 70(5):153–63.

Rosenbaum, Paul R. 2010. *Design of Observational Studies*. Springer.

Ross, Joel, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. 2010. “Who Are the Turkers? Worker Demographics in Amazon Mechanical Turk.” *Chi Ea 2010* (July 2016):2863–72.

Tyler, T., J. Fagan, and A. Geller. 2014. “Street Stops Police Legitimacy: Teachable Moments in Young Urban Men’s Legal Socialization.” *Journal of Empirical Legal Studies* 11(14):751–85.

Tyler, Tom R. 2017. “Procedural Justice and Policing: A Rush to Judgment?” *Annual Review of Law and Social Science* 13:29–53.

Tyler, Tom R. and Jonathan Jackson. 2014. “Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation, and Engagement.” *Psychology, Public Policy, and Law* 20(1):78–95.

Walters, G. D. 2017. “Beyond Dustbowl Empiricism: The Need for Theory in Recidivism Prediction Research and Its Potential Realization in Causal Mediation Analysis.” *Criminal Justice and Behavior* 44(1):40–58.

Walters, Glenn D. and Jon T. Mandracchia. 2017. “Testing Criminological Theory through Causal Mediation Analysis: Current Status and Future Directions.” *Journal of Criminal Justice* 49:53–64.

Worden, Robert E. and Sarah J. McLean. 2017. *Mirage of Police Reform: Procedural Justice and Police Legitimacy*. University of California Press.

Yamamoto, Teppei. 2014. “Identification and Estimation of Causal Mediation Effects with Treatment Noncompliance.” *Draft* 35.

Yesberg, Julia and Ben Bradford. 2018. “Affect and Trust as Predictors of Public Support for Armed Police : Evidence from London.” *Policing and Society* In Press.

Epilogue

This thesis started with Nagin and Telep's (2017) observation, which correctly highlighted the scarcity of established causal connections in the procedural justice literature. In response to their statement, the contributions of this thesis were twofold. First, it provided a toolkit for researchers who want to test causal mechanisms. This was done by reviewing the literature on causal mediation analysis, which has been a burgeoning field in epidemiology and biostatistics. Second, it proposed a comprehensive model of procedural justice and assessed the causal connections between the various constructs. The causal evidence collected in the four papers provides one of the first credible indications that many of the expected mediated effects are not mere associations but causal pathways.

A basic message of this thesis is therefore that procedurally just policing works. When the police give people a voice, explain their actions, respect their boundaries, and treat them with dignity and respect, citizens are more likely to find police behaviour morally appropriate and recognise their authority to dictate appropriate behaviour. Police legitimacy is important because it is the key to achieving certain societally desirable outcomes, such as cooperation with the police and compliance with the law – and crucially, certain aspects of legitimacy are more important in influencing one or the other. Moreover, procedurally just treatment elevates an individual's personal sense of power, which is capable of transmitting the effects of procedural justice to certain parts of police and legal legitimacy. I recognise that these four papers are only the first steps in testing and identifying causal mechanisms in the procedural justice literature, and I am looking forward to continuing this work in the years to come.