New Directions of Research in Fairness and Legal Authority: A Focus on Causal Mechanisms

Invited book chapter: Social Psychology and Justice (Frontiers of Social Psychology), edited by E. Allan Lind. Oxon: Routledge.

Jonathan Jackson*, Department of Methodology, LSE
Krisztián Pósch, Department of Methodology, LSE

*Corresponding author: Jonathan Jackson, Department of Methodology, LSE, Houghton Street, London WC2A 2AE, United Kingdom. Tel: +44 207 9557652. Email: j.p.jackson@lse.ac.uk
ORCID: 0000-0003-2426-2219

Abstract: We consider three new directions of research into police-citizen authority relations: (a) the subjectivity of fairness perception, (b) the mechanisms linking procedural justice to legitimacy, and (c) statistical methods to estimate causal mechanisms. First, police fairness may be a subjective experience and perception motivated by a range of individual and environmental factors: we outline a motivation cognition framework of fairness perception (Barclay et al., 2017) that revolves around people 'reading' the dynamics of an encounter with the police in a way that is shaped by instrumental, relational and moral motives, with directional and non-directional goals. Second, the causal effect of procedural justice on legitimacy may partly be transmitted by social identity and personal sense of power and autonomy. Third, causal mediation analysis is complex and we show how to decompose the average treatment effect in a way that allows for the estimation of causally mediating and moderating effects, provided that certain causal identifying assumptions are satisfied (VanderWeele, 2014).

Key words: legitimacy, police, mediation, moderation, social identity, procedural justice, sense of power.

Acknowledgements: the analysis and writing of this paper was funded by the United Kingdom's Economic and Social Research Council:

- ES/R011397/1. *From Coercion to Consent: Social Identity, Legitimacy, and a Process Model of Police Procedural Justice.*

Decades of research into procedural fairness and criminal justice have shown the importance of procedural justice and injustice in the generation, maintenance and erosion of people's perceptions of the legitimacy of legal institutions (Tyler, 2006a, 2006b; Sunshine & Tyler, 2003; Tyler & Fagan, 2008). While most of this work has addressed people's relationship to the police using city-wide or nationally representative sample surveys (for a review of the international literature see Jackson, 2018), the past few years have seen an increasing amount of experimental work, with researchers using online and offline textual and video vignettes of police-citizen encounters (e.g. Solomon, 2019; Pósch, 2019; Trinkner et al. 2019; Radburn et al., 2018; Lowrey et al., 2016; Maguire et al., 2016; Barkworth & Murphy, 2015) and randomized controlled trials to estimate causes and consequences of police and legal legitimacy (e.g. Pósch et al., 2019; MacQueen & Bradford, 2016; Mazerolle et al., 2013).

The evidence from this expanding body of work supports two points that are central to the current contribution. On the one hand, citizens place more importance in their direct and indirect encounters with the police on whether officers treat people with respect and dignity (the quality of treatment) and make neutral, unbiased decisions (the quality of decision-making) than whether they deliver satisfactory outcomes, are effective in the fight against crime, and/or allocate outcomes fairly across social groups—albeit that respecting the limits of one's rightful authority may also be an important consideration (Huq et al., 2017; Trinkner et al., 2018). On the other hand, police officers create legitimacy through their interactions with the public (Geller & Fagan, 2019) and people update their beliefs about the legitimacy of the police—perceptions of institutional appropriateness (right to power) and entitlement (authority to govern)—largely on the basis of fair treatment and fair decision-making from the actors who embody the institution. Because procedural justice is more closely tied to perceptions of police legitimacy than the effectiveness of officers to fight crime and satisfaction with outcomes—and because legitimacy is typically a stronger predictor of legal compliance than more instrumental concerns about the likelihood and cost of punishment—scholars have called for consensual modes of policing based on restrained use of authority over more pro-active, coercive modes of policing (Tyler et al. 2015; Bradford et al., 2013).

The behaviours that foster perceptions of police procedural justice are generally classified under the headings of (i) decision-making (giving people a voice, making impartial decisions that are free from bias, and explaining the reasoning behind decisions) and (ii) interpersonal treatment (behaving in respectful ways, showing care and concern, being honest, and conveying trustworthy motives). Procedural justice is an important cultural norm regarding the appropriate exercise of police authority, and procedurally (in)just treatment and decision-making shapes people's belief that the police is a normatively appropriate institution that has the right to dictate appropriate behaviour (Tyler, 2006a, 2006b). In turn, legitimacy seems to help strengthen people's normative motivations to comply with the law (de Puiseau et al., 2019; Trinkner et al., 2018; Tyler & Jackson, 2014; Murphy et al., 2014). There is also some early evidence that some of this statistical effect of procedural justice on legitimacy is mediated by social identity, whereby people are motivated to legitimate authority figures of groups that they identify with (Bradford et al., 2014).

Yet, despite the growing number of experimental studies testing procedural justice theory, there has been an over-reliance on correlational cross-sectional methodology (Nagin & Telep, 2017a, 2017b; Tyler, 2017). The arrow of causality is generally assumed to flow from perceptions of procedural justice to perceptions of legitimacy. But Nagin & Telep (2017) have recently raised the possibility that the relationship between procedural justice and legitimacy is bi-directional. Prior levels of legitimacy may shape how people make sense of the dynamics of direct and indirect police-citizen encounters—in the words of Trinkner et al. (2019: 3): '…one could imagine a case in which individuals who believe police officers are legitimate authorities would be more likely to judge police behavior as procedurally fair than individuals who believe the police are illegitimate.' By motivating people to see police as fair, high levels of legitimacy could render fairness perceptions difficult to change, with people *giving the police the benefit of the doubt* in the face of ambiguity.

In this chapter we discuss three new directions of research into police-citizen authority relations: (a) the subjectivity of fairness perception, (b) the mechanisms linking procedural justice to legitimacy, and (c) statistical methods to estimate causal mechanisms. Our goals are threefold. The first is to discuss the idea that fairness is a subjective experience/perception motivated by a range of individual and environmental factors. We outline a motivation cognition framework of fairness perception (Barclay et al., 2017) that revolves around people 'reading' the dynamics of an encounter

with the police in a way that is shaped by instrumental, relational and moral motives, with directional and non-directional goals. The idea is that people attend to, and process, information that helps them to reach their desired conclusion regarding fairness, especially when those dynamics are ambiguous, i.e. where it is not an immediately obvious instance of fair or unfair treatment and/or decision-making. We speculate about the possibility of 'temporal stickiness' of fairness perceptions. We discuss (i) fairness heuristic theory (Lind et al., 2001), where people initially form general justice judgments regarding how fair they think the police are, and then, once a relatively stable fairness heuristic is 'set', it takes a particularly vivid and unexpectedly fair or unfair encounter to change one's general impression of fairness. We also discuss (ii) the idea that the deference part of the legitimacy construct means letting the police dictate appropriate behaviour between citizens and officers, meaning that people who see the police as legitimate are more likely to side with them when it comes to ambiguous (un)fairness.

The second goal of this chapter is to consider another under-researched issue—namely, the causal mechanism(s) that transmit and/or modify the effect of procedural justice onto legitimacy. Research tends to address the direct effect of both the experience of procedural (in)justice and general perceptions of police procedural (un)fairness on legitimacy. This is a direct (i.e. unmediated or moderated) effect that is about respecting social norms regarding the appropriate exercise of authority. Legitimacy is the belief that the institution is appropriate coupled with an internalised obligation to obey, and procedural justice a key core legitimating norm regarding how power should be wielded. There has, however, been some non-experimental work on the idea that social identity is a (partial) mediator of this effect (Bradford et al., 2017; Bradford, Murphy, & Jackson, 2014; Radburn & Stott, 2018). This account is in line with the group engagement model (Blader and Tyler 2009; Tyler and Blader 2003) and it argues that procedural justice sends a message of status and value to the justice recipient, which in turn encourages the justice recipient to identify with the social group that the police represent, thereby strengthening the legitimacy of authority figures of the group that one merges one's identity with.

We consider the idea that social identity and personal sense of power/autonomy are two (not mutually exclusive) causal mechanisms that transmit some of the effect of the experience of procedural (in)justice (and more general perceptions of procedural (un)fairness) onto legitimacy. Personal sense of power refers to one's subjective belief regarding the ability to influence the police during potential future encounters, where a strong sense of personal power reflects the belief that police have little power over oneself (i.e. one has considerable autonomy) and a weak sense of personal power reflects the belief that police have a lot of power over oneself (i.e. one has little autonomy). Personal sense of power and autonomy may shape legitimacy if some basic level of personal autonomy is central to the psychological phenomenon of believing power is rightfully held and willingly accepting the authority of another to dictate appropriate behaviour. Legitimacy defines a consensual rather than a coercive relationship (Anderson et al., 2015) and procedural justice signals less of an assertive and/or aggressive stance that conveys a power-holder/subordinate relationship, and more of what Mentovich (2012: 15) calls:

> "…a community, equity prioritising cue…By emphasizing shared values, goals and equal entitlements, procedural justice may conceal the power structure of a given community in favor of a more communal perspective, resting on a perception of equality."

The enactment of procedural justice in a personal encounter with the police may thus send not just a symbolic message that the citizen has status and value within hierarchical group settings, but also that the officer respects the autonomy of the citizen. These two messages signalled by procedural justice (of status/value and of more equal power relations) may, in turn, help to encourage the justice recipient to view the institution that the officer embodies as legitimate.

The third goal is to highlight a burgeoning subfield of the causal inference literature that has the potential to estimate causal mechanisms (Keele et al., 2015; Pósch, 2019; VanderWeele, 2015; VanderWeele & Richardson, 2012). Experimental studies commonly only estimate the effect of a randomised treatment on a certain outcome, and while this simple setup can be useful to evaluate the effectiveness of certain initiatives, it does fall short in explaining why and how the treatment reached or failed to reach its goal. Our empirical illustration of causal mediation and causal interaction analyses focuses on social identity as a mediator and moderator of the effect of a treatment (showing pictures of police officers vs civilians) on different aspects of police legitimacy. We show how to decompose the

average treatment effect in a way that allows for the estimation of causally mediating and moderating effects, provided of course that certain causal identifying assumptions are satisfied (VanderWeele, 2014). To round up the discussion of causal mechanisms, we revisit two potentially confounding factors emphasised by Nagin and Telep (2017)—'third common causes' and 'reverse causality'—and suggest a few ways to tackle them.

The rest of the chapter proceeds as follows. In the next section we provide a brief criminological introduction into the nature of policing and the complexities of police-citizen encounters. Drawing then on Barclay et al.'s (2018) recent review, we outline a motivated cognition approach to police procedural fairness. After this we discuss two reasons why there may be 'temporal stickness' to people's perceptions of police fairness, and to motivate why perceptions of fairness may be difficult to surprisingly stable in the face of evidence, we consider (a) Lind's fairness heuristic theory and (b) the idea that the deference part of legitimacy shapes non-directional and directional goals. We then turn to the issue of causal mechanisms. While we do not (yet) have the data to test the outlined ideas, we do have a parallel dataset that provides a good example for an illustration of appropriate statistical analysis.

Everyday policing

It was Bittner (Bittner, 2005, Bittner, 1967, Bittner and Bish, 1975) who first pointed out that the police are in the unique position of being able to impose immediate solutions to the constant flow of small conflicts, irregularities and problems that need to be dealt with in society. Actual crime-fighting is only a fraction of what the police do—whenever there is '*something-that-ought-not-to-be-happening-and-about-which-someone-had-better-do-something-now!*' (Bittner 2005, p.161) it is a task for the police. Because police officers develop skills for handling complex and ambiguous social environments by learning from colleagues and practice, they are better understood as craftsman rather than bureaucrats.

The relationship between the police and the public is also partly characterised by efforts by the police to reaffirm their power and their ability to exercise this power and authority to varying degrees (Waddington, 1996a, 1996b). What characterises their activity is their authority and ability to coerce using force, where the ability to threaten force is often enough to control a situation (Skolnick, 1966). The police dramatise the appearance of control, manage the information available to judge their success, and seek to establish an appearance of unity of purpose. Because they desire public confidence, but they cannot show that they are successful with their mission of public control, they look to establish their authority and legitimacy by employing the sort of interpersonal communication strategies that Goffman (Goffman, 1958, Goffman, 1961, Goffman, 1967) so classically described (Manning, 1977).

There are a variety of different types of police-citizen encounters: there are involuntary foot and road stops; there are citizens reporting a crime; there are third-party observers observing an officer patrolling the streets; and so on and so forth. Different dynamics can emerge and develop, even in short encounters, depending on the context, the behaviour of the citizens, and the behaviour of the officers. Take a police stop. In such an instance, in the words of Nagin and Telep (2017: 3): 'Hostile and disrespectful behavior on the part of authority figures is likely to provoke an angry response or only grudging compliance with their orders.' Equally, hostile and disrespectful behaviour on the part of citizens is likely to provoke a more assertive show of authority from officers. Indeed, the levels of hostility, respect, compliance and resistance could ebb and flow over the course of a single encounter—e.g. an officer could encounter an initially friendly and compliant citizen, but the citizen could then turn less compliant to the officers requests, leading the officer to assert her authority in a stronger manner, with the citizen grudgingly backing down as a result.

There can, as a result, be considerable ambiguity in the culpability, appropriateness and deservingness of police and citizen behaviour in a given encounter. To explore how different people make sense of policing, Waddington et al. (2018) presented recordings of four different policing events to people taking part in a series of focus groups. There was a group of officers turning up at the scene of a robbery of an elderly man in his home; there was a group of officers stopping a suspected stolen car on the motorway; there was a suspected car theft in a supermarket car park; and there was a violent arrest outside a nightclub. The discussions that emerged from these focus group brought to life the fact that officers are both protectors and regulators; that discretion is fundamental to policing; and that the speed with which officers make sense of ambiguous and complex situations and apply legal categories can be dizzying to those of us who do not do this job. Waddington et al. (2018) also showed that the same encounter can be seen by different observers in quite different ways. While the evaluative criteria

used by research participants were largely the same—people in the focus groups generally focused on (a) the procedural justice displayed by officers and (b) whether officers seemed to respect the limits of their rightful authority, particularly in the case of regulatory encounters involving violence or the potential for violence—there was striking variation in the application of these criteria. This variation often seemed to be linked to the degree to which research participants identified either with the citizen or with the officer. For instance, while an officer giving a second fine for the same traffic rules violation seemed reasonable to members of a neighbourhood watch, young offenders serving their community service believed that the officer was just 'taking the piss' (Waddington et al. 2018: 61).

Why might prior levels of fairness and/or legitimacy play a role in fairness perception?

Procedural justice theory (PJT) is an increasingly popular theoretical approach through which to view police-community relations (e.g., Tyler 2006a, 2006b, 2009; Bradford et al., 2014; Tyler et al. 2015; Tyler, 2017; cf. Weisburd et al., 2019). With its central focus on what generates consensual rather than coercive relationships, PJT resonates strongly with the ideology of policing by consent. Premised on the idea that most people obey the criminal law most of the time because they think it is the 'right thing' to do so, and not simply because it is in their own best interests to do so, core tenets of PJT – e.g. the idea that 'fair' policing builds legitimacy and the idea that legitimacy enhances consent-based relationships between police and public – have become widely among some academics and policy-makers (e.g., President's Task Force on 21st Century Policing, 2015; Hough et al., 2013). PJT makes four main predictions:

(a) that it is the *style* of social interaction and the *neutrality* of decision-making in encounters between individuals and justice officials that are crucial in shaping behaviour;
(b) that the social bonds between individuals and institutions are strengthened when authorities make fair and neutral decisions, and when people are treated in ways that are recognised to be fair, respectful and legal – and not based on bias and stereotypes;
(c) that out of these social bonds comes a sense that legal authorities are legitimate – that the police and courts have the right to power, the right to dictate appropriate behaviour, and are morally justified in expecting cooperation and compliance; and,
(d) that legitimacy promotes normative modes of compliance and cooperation that are both more stable and more sustainable in the long run than models of policy based on deterrence, sanction and fear of punishment.

PJT specifies the direction of causation as going from perceived procedural fairness to legitimacy. Yet, as outlined earlier, legitimacy could plausibly shape fairness perception. Fairness is in the *eye of the beholder* and different people might come to different conclusions about the same police-citizen interaction because their perceptions are influenced by subjective frames emergent from different previous experience. In this section we consider the idea that motivated cognition (Barclay et al., 2017) provides a framework through which we can develop a programme of research into police fairness perceptions.

A motivated cognition approach to perceived police fairness

According to Barclay et al.'s (2018) review, there are instrumental, relational and moral reasons why people care about fairness, and these different (instrumental, relational and moral) motives can differentially focus attention on fairness-relevant information to either highlight or ignore unfairness. Non-directional goals are focused on accuracy and reaching the right conclusion—when activated, people make more effort to attend to and process information, use more complex rules, and so forth. Directional goals encourage people to use strategies (of attending to and processing information) that will help them reach a desired conclusion—they shape the beliefs and strategies that are used to guide and support conclusions (albeit with the constraint that the conclusions can be rationally justified). Non-directional and directional goals can operate at the same time, e.g. people could put more effort in, with the goal of reaching a desired conclusion. As Barclay et al. (2018: 13) also state: 'These goals may also activate different emotions; instrumental motives may activate envy, relational motives may enhance loneliness, whereas moral motives can spark moral outrage.'

What is the relevance of motivated cognition to police fairness perceptions? Instrumental motives start with the idea that when police act and make decisions in fair ways, this can help people feel that they have control over—and can more generally expect—beneficial outcomes (see, for example, Thibaut and Walker, 1975). Respectful treatment is more pleasant than disrespectful treatment, but it also signals to the justice recipient that she may receive a positive outcome from the interaction, and having a voice in the interaction helps her feel that she can avoid something negative happening, as does neutral and unbiased decision-making from the officer.

How might instrumental motives shape the experience and perception of direct (personal) and indirect (third-party observer) police-citizen encounters. On the one hand, instrumental motives may shape someone's behaviour in a personal (direct) encounter with an officer; she wants a positive outcome so she acts respectfully, is compliant, and so forth; and she pays self-interested attention to cues of fairness from the officer to ascertain whether she will benefit or avoid a negative outcome. On the other hand, instrumental motives may shape the perceptions of a third-party observer of a police-citizen interaction. As Barclay et al. (2018: 11) state: '…observing unfairness can spark self-interested concerns because observers want to avoid comparable treatment and/or have internalised the harm caused to another.' Instrumental concerns could motivate people in a non-directional way. Someone might put more effort into reaching as accurate a judgement as possible regarding the fairness of the officer to, for instance, avoid comparable treatment. Instrumental concerns could also motivate fairness perceptions in self-interested directional ways. One might be more likely to perceive unfairness when one identifies with the injustice recipient (Brockner & Greenberg, 1990) and feels that one's own outcomes may be threatened in a similar encounter (Chaikin & Darley, 1973).

Relational motives focus on the symbolic messages that fairness (particularly procedural justice) send to justice recipients regarding their status, value and inclusion/exclusion within the group that authority figures represent (Lind & Tyler, 1988; Tyler & Lind, 1992). The superordinate group is plausibly 'law-abiding members of the community' (Bradford et al., 2014) and relational motives could shape a person's behaviour in a direct encounter with an officer: she wants to receive positive messages of her standing in the community, so she treats the officer with respect and dignity, shows deference to the uniform, and generally tries to signal to the officer that she is respectable. She would also pay attention to how the officer is behaving because of the relational issues at stake. As a third-party observer she might be motivated to perceive unfairness when she identifies with an injustice recipient—she internalises the negative relational message that is being sent to the other individual. Conversely, she may be more motivated to perceive fairness— she tries to avoid internalising the negative relational message so she views the injustice recipient as outside of her "scope of justice" (Opotow, 1990), blames the victim and identifies with the officer.

Moral motives are about internalised values and norms. People care about fairness for moral reasons—they believe that the police *should* exercise authority in fair ways because that is the right thing to do. Procedural justice is a strong societal value dictating how authority should be exercised (Tyler, 2006a, 2006b; Hough et al., 2010; Stanko et al., 2012) and there may also be a widely-shared belief that citizens should, *in reciprocation,* treat police with respect and deference. In a direct encounter, people with strong moral motives might be respectful and deferent because (a) that is the right thing to do and (b) it encourages reciprocity. People may also pay extra attention to the quality of treatment and decision-making of the officer. As a third-party observer, the extent to which individuals internalise these norms may shape directional goals regarding how they make sense of the fairness of the officer. If one observer believes that citizens should respect the police, any deviation from the expected level of deference from the (in)justice recipient might lead an observer to 'side' with an officer and believe that the citizen got what she deserved. Conversely, the belief in a just world could be threatened when one sees unfairness, leading to moral outrage and a desire to restore justice (Skarlicki et al., 2015).

On the 'temporal stickiness' of perceived police fairness

In addition to the possibility that there are instrumental, relational and moral motives to care about fairness that shape non-directional and directional goals of judgement, there may also be some kind of 'temporal stickiness' at play in perceived police fairness. Once an individual has arrived at an initial judgement regarding the general fairness of police, this judgment may be difficult to change. One

possibility is nicely elaborated in Allan Lind's fairness heuristic theory (Lind et al., 2001; Van den Bos et al., 2001). Another, not mutually exclusive, explanation revolves the deference part of legitimacy.

According to Lind (2001), people use fairness (not just procedural fairness, but also distributive and interactional fairness) as a short-cut to understanding the positive or negative nature of their relationship to a social group. Believing that one has been treated fairly encourages one to shift one's perspective from the individual to the group. The social group could be something like an individual's relationship to the organisation within which she works. The positive or negative relationship manifests itself in terms of levels of trust, acceptance of authority and rules, identification and the willingness to proactively cooperate. People need to quickly make decisions in the presence of a possible cost to cooperation and risk of exploitation, and the fairness heuristic helps them to do this. As Lind (2001: 65-66) says:

> "I am suggesting that people use overall impressions of fair treatments as a surrogate for interpersonal trust; that they refer to their impressions of fair or unfair treatment as they process the requests, demands, and potential obligations that are so much a part of social and organizational life. People use fairness judgments in much the same way that they would refer to feelings of trust—if they had an independent basis for forming trust—to decide how to react to demands in a long-standing personal relationship. It is important to note that I am not suggesting that people are generally conscious of either the social contract implicit in the foregoing discussion of the fundamental social dilemma or the solution of the dilemma through the use of fairness judgements. I am suggesting instead that over the course of socialization and especially in the course of learning about the potential costs and benefits of associating with and identifying with others, we come to use our impressions of fairness as a guide and to regulate our investment and involvement in various relationships to match the level of fairness that we experience."

At the beginning of a group relationship, when someone first encounters authority figures, for instance, one begins a "judgment phase" in which one attends to signals of procedural, distributive and interactional fairness. Once a general justice judgement is formed, people shift to the "use phase", in which the now relatively stable judgement guides not only trust, behaviour, identification etc, but also procedural, distributive and interactional fairness judgements. Importantly for the current discussion, the justice judgement is likely to be revised only when the relationship clearly seems to be changing or when an unexpectantly vivid, fair or unfair thing happens. As Lind (2001: 70-71) describes: "I am suggesting that once a general justice judgment, it will be assumed to be accurate, and any incoming information relevant to the fairness of treatment will be reinterpreted and assimilated to be congruent with the existing general fairness judgement."

While fairness heuristic theory is clearly relevant to organizational group settings, how applicable might it be to police-citizen relations? Lind et al. (2001: 190) states that: "…the fairness heuristic is activated when people expect substantial interaction with a person or group (and thus run the risk of exploitation) or when they identify with a group or relationship (and therefore can suffer identity loss from exclusion)." While only a relatively small proportion of people in most societies may have 'substantial interaction' with the police, identification with a superordinate law-abiding member of community group seems on the face-of-it to be meaningful for most people. The police can be considered condensation symbols of the state enforcing and maintaining order to the surrounding social world (Loader, 2008) and the police patrol the boundaries of respectability (Waddington, 1999). People may therefore be motivated to have a stable and positive relationship with the police and the demands that the police make on people's behaviour. Thus, even though most people do not have significant early interactions with the police—they are socialised into the idea of the police and law through parents, teachers, the media, their peers, etc—they may nevertheless form a general justice judgement about the police that does not change very much, using the relatively stable heuristic to understand their own personal relationship with the police.

Legitimacy may play a similar role in rendering fairness perceptions 'sticky' and unlikely to change. As a perception of the rightfulness of the institution that officers embody—a structural quality that gives the individuals who hold the officer power and authority—legitimacy is typically defined and measured along two connected dimensions. The first is the belief that the institution is normatively

appropriate, i.e. moral, just and appropriate within the given societal context. The second is the belief that the institution has the right to dictate appropriate behaviour, where individuals internalise the moral value that they should *obey the police because they're the police.* Legitimacy means ceding the right of a power-holder to dictate appropriate behaviour—not only of citizens and police, but also the appropriate interpersonal dynamics between citizens and officers—and deference is about an individual respectfully submitting to the judgement of a power-holder. When it comes to people's own experience, if people view the police as legitimate they will defer to them more readily; compliant behaviour from citizens reduces the propensity of officers to be more aggressive. With indirect experience with the police (as a third-party observer), it may be that legitimacy leads people to *give the officer(s) the benefit of the doubt* in the event of ambiguous dynamics and ignore or overlook unfairness.

For example, it may not be abundantly clear to the third-party who instigated a turning point in which an officer moved from more respectful (consensual) language to more aggressive (coercive) language. An observer who viewed the police as a legitimate institution might reason that the officer was correct to act in the way she acted; deference means allowing the officer to decide how all those involved should rightfully behave (e.g. the movement to more assertive language was justified). In terms of non-directional (accuracy) goals, deference may mean that one makes little effort in attending to and processing information about fairness, assuming without much cognitive effort that the officer was fair. In terms of directional goals, deference may mean that one processes information in a way that comes down on the side of the officer. Perhaps one decides that the citizen was to blame for the dynamics of an interaction that meant that the police went in the direction of more assertive and aggressive behaviour?[1] Just as the fairness heuristic might mean that it takes an unexpectedly fair or unfair encounter to *move around* someone's general sense of police fairness, high levels of legitimacy might render people uncritical and/or unquestioning of police behaviour, whether specific or general.

At the same time there is a fair amount of correlational criminological evidence regarding the asymmetry of impact of police-citizen interactions. In particular, negatively received encounters tend to correlated with a good deal lower levels of perceived trustworthiness, confidence and legitimacy, while positively received encounters tend to correlated with at best only slightly higher levels of perceived trustworthiness, confidence and legitimacy (Skogan, 2006; Bradford et al., 2009; Jackson et al., 2013). Consistent with the idea that perceptions of fairness can change as a result of negative encounters with the police (which may be more likely to stimulate a shift in goals), a motivated cognition perspective suggests that fairness perceptions could change more readily, as Barclay et al. (2018: 8) outline:

> '…motivated reasoning would suggest that a single shift in goals (e.g. from accuracy to feeling good) can just as easily lead to "phase shifting." Applying a motivated perspective can allow predictions about the extent to which the new heuristic would be bound by both accuracy and directional goals, thereby providing a more fine-grained approach to the phase shifting process.'

It is for future research to provide empirical evidence on what shapes fairness perceptions and how responsive trust, confidence and legitimacy is to people's direct and indirect experience.

Identity and power as mechanisms for the procedural justice-legitimacy relationship

The second discussion point of the current chapter revolves around the social psychological mechanisms linking procedural justice to legitimacy. The standard approach to interpreting the impact of the experience of procedural justice and/or injustice on legitimacy focuses on norms of conduct and the normative appropriateness of an institution. Most criminological work looks only at the direct relationship from procedural justice to legitimacy, whereby people draw on information about procedural justice to infer that the institution that she represents is normatively appropriate. Procedural

[1] In addition to the effect that legitimacy may have on fairness perceptions in the context of a single interaction, prior levels of legitimacy may also moderate the impact of a given encounter on subsequent beliefs about general police procedural fairness. More specifically, deference may encourage citizens to differentiate between one unjust police-citizen encounter and the general pattern of police behaviour. Even in the face of an unfair encounter, one gives the institution *the benefit of the doubt*, by discounting this particular officer's behaviour as atypical and not representative of a general trend.

justice is a core societal value dictating the appropriate use of police authority, and when an officer is seen to act in procedurally fair ways, this enhances and/or maintains legitimacy in the eyes of citizens.

However, it may be that social identity mediates some of the effect of procedural justice on legitimacy. There is a good deal of evidence that people care about procedural justice, in part because of the relational message conveyed by fair process (Blader & Tyler, 2009). As Blader & Tyler (2015 [oxford handbook of justice in the workplace]) note: 'Procedural justice conveys a positive message to justice recipients about their relationship with the entity enacting justice, whereas procedural injustice conveys a negative message about that relationship.' When an officer acts fairly in terms of decision-making processes and interpersonal treatment, this sends a message of status and value to the justice recipient, encouraging the justice recipient to identify with the social group that the police represent (Tyler & Blader, 2002, 2003)—as, perhaps, a law-abiding citizen (Bradford et al., 2014). When one merges one's self-concept with a group with authority figures that make demands on one's behaviour, one is motivated to legitimate those authority figures to help maintain one's positive connection to the group.

It may be that there is one more psychological mechanism linking procedural justice to legitimacy. It was Mentovich (2012) who first looked at the effect of the enactment of procedural fairness by high-power authorities on people's psychological experience of power. Procedural fairness has a relational effect via the signalling of status and value, but it also has a relational effect via a "…community, equity prioritising cue…By emphasizing shared values, goals and equal entitlements, procedural justice may conceal the power structure of a given community in favor of a more communal perspective, resting on a perception of equality" (p. 15). By activating a communal rather than hierarchical schema, the experience of procedural justice reduces perceptions of differential power. This works through the heightened sense of control that procedural justice instills and the greater sense of in-group status.

We define personal sense of power and autonomy as the subjective belief about one's ability to have some sort of influence and agency in some future (hypothetical and actual) encounter with the police (Posch, 2019). Why might personal sense of power and autonomy of behalf of individuals (in relation to the police) shape people's perceptions of police legitimacy? Legitimacy transforms someone's orientation towards the police away from coercive power (to arrest, to use violence, and so forth) towards a sense of willing and active consent to the demands that the police make over one's freedom. This consent is most obviously with respect to enforcing the law, but it also refers to the police's implicit and explicit claim to generally dictate appropriate behaviour. Legitimacy has been linked to greater empowerment of police powers (Sunshine & Tyler, 2003), the belief that the police do not view oneself as an object of suspicion (Tyler et al, 2015), and an acceptable of the right of the police to monopolize violence (Jackson et al., 2014).

Criminologists talk about procedural justice policing being a consensual rather than a coercive mode of policing, where officers police as part of the community not against the community. Part of this dynamic may be the downplaying and/or concealing of the imbalance of power through procedurally fair treatment and decision-making. For example, making citizens feel that they are not an object of suspicion may encourage consent. For it to be true legitimacy, consent must be an active, autonomous choice, and there may be an important 'sweet spot' in terms of how much coercive power one feels the police to have over oneself (and the extent to which one believes that the police look at oneself as an object to control). If one feels that one has little power, it may be difficult to willingly consent.

Statistical tools to estimate causal mechanisms

To empirically test whether social identity *and* personal sense of power mediate some of the effect of procedural justice to legitimacy, one needs to estimate causal mediation. The focus on causal mechanisms in the social sciences is relatively novel, with Sampson et al. (2013) view causal mechanisms being one of the largest challenges of translating research findings to actual policy. They argue that while randomised controlled trials (RCTs) can establish strong internal validity and can derive unbiased average treatment effects from an intervention to the outcome, they come short in explaining *why* and *how* the intervention was (in)effective. This makes it challenging to assess the transferability and generalisability of the findings. In response, Sampson et al. (2013) recommend a greater focus on causal mechanisms for at least three reasons: (1) they can help disentangle real causes

from confounders, (2) they can inform policymakers why a certain route needs to be taken during the implementation, and (3) they permit comparison of alternative pathways by juxtaposing mechanisms. Others have also advocated focussing on causal mechanisms for methodological (Matsueda 2017) and theoretical (Kirk and Wakefield 2018) reasons.

While it is difficult to come up with a single definition of causal mechanisms because of the multitude of philosophical approaches to causation (Beebee, Hitchcock, and Menzies 2009), Hedström and Ylikoski (2010) put forward four notable characteristics shared by most definitions, where causal mechanisms:

1. …are construed by the causal effect or phenomenon that produced them (e.g., perceived legitimacy of the police is produced by the perception of procedural justice);
2. …are causal notions which correspond to the process that creates the effect of interest (e.g., procedural justice produces an effect on police legitimacy);
3. …establish structure, which makes the black box of causality transparent (e.g., initial views of procedural justice influence later views of procedural justice via the deferential aspect of police legitimacy); and,
4. …create a hierarchy, where certain effects precede other effects (e.g., procedural justice affects police legitimacy but not the other way around).

The most confusing aspect of causal mechanisms is that they possess duel properties: they simultaneously describe the causal process and *how* certain effects arise (i.e., 'how does perceptions of procedural fairness change over time?') and produce the subsequent effect explaining *why* the intervention works (i.e., 'why does procedural justice influence legitimacy?').

There are several methodological approaches, including focus groups and interviews (Haberman 2016; MacQueen and Bradford 2017), mixed methods (Johnson, Russo, and Schoonenboom 2017; Weller and Barnes 2016), process tracing (Fairfield and Charman 2017; Saylor 2018), and network analysis (Papachristos et al. 2012; Papachristos, Wildeman, and Roberto 2015), but they all have major limitations. For example, qualitative research has been primarily conducted *ex post facto*, addressing only cases that have turned out to be notable or startling (Carson and Wellman 2017; Hassell and Lovell 2015), while network analysis is still in its infancy and generally incapable of deriving causally interpretable effects (Ogburn et al. 2017; VanderWeele, Ogburn, and Tchetgen Tchetgen 2012).

We advocate using a family of methods that has been largely developed in the past decade: the analysis of causal mediation and causal interactions (e.g., Imai et al. 2011; Keele 2015; VanderWeele 2015). These methods rely on the potential outcome framework to describe causal effects by decomposing the average treatment effect and providing insight into the extent to which the treatment effect is mediated, is due to an interaction, or to do mediation and interaction together (Vanderweele 2014). We consider a randomised experiment to mirror the virtual reality experiments we are planning to conduct to address the two gaps in the literature we have described in this chapter. We briefly discuss causal mediation analysis and causal interactions and effect heterogeneity. Then we demonstrate the use of these methods with an empirical example. Finally, we address some common criticisms regarding 'reverse causality' and 'third common causes'.

Causal mediation analysis

The intuition behind causal mediation analysis is similar to the mediation analysis routinely used in structural equation modelling (Baron and Kenny 1986), postulating that the average treatment effect (or the total effect) can be decomposed into direct and indirect effects if the causal identification assumptions are met. The direct effect is the unmediated effect of the treatment on the outcome and the indirect effect is the part of the treatment effect that is transmitted by an intermediate variable towards the outcome. The traditional Baron and Kenny (1986) approach to mediation analysis suffers from several limitations, including (1) the assumption of unit-level effect homogeneity (the treatment effects need to be the same for each individual), (2) the linearity assumption (the additivity of the effects is not possible in non-linear models) and (3) the no-interaction assumption (the decomposition breaks down in the presence of a treatment-mediator interaction that affects the outcome). Causal mediation analysis can overcome these limitations by offering nonparametric identification of the direct and indirect effects

that can effortlessly integrate the interaction. This permits more flexible modelling (Imai et al. 2011; Imai, Keele, and Tingley 2010).

For the causal identification of the direct and indirect effects the sequential ignorability assumption needs to be satisfied (Imai et al. 2010; Pearl 2001). This is a set of no-unmeasured confounding assumptions (akin to matching) that states that, controlling for influential pre-treatment covariates, there is no unmeasured confounding for:

1. The treatment-outcome relationship
2. The mediator-outcome relationship also controlling for the treatment
3. The treatment-mediator relationship
   and also,
4. There is no other mediator that has been affected by the treatment

From these assumptions, the first and third are satisfied in randomised experiments. However, there is no easy way to rule out that an unmeasured confounder might affect the relationship between the mediator and the outcome (second assumption). The fourth assumption can also be untenable in many instances as often multiple mediators are presumed to mediate the effect of the treatment on the outcome. This includes the procedural justice literature, wherein several studies into police legitimacy have been presumed to have two aspects: normative alignment and duty to obey (Jackson et al., 2012; 2013).

The second assumption of the sequential ignorability assumption could be easily remedied by the manipulation of the mediator. Although there are special experimental designs, such as the crossover (encouragement) and parallel (encouragement) designs (Imai, Tingley, and Yamamoto 2013; Pirlott and Mackinnon 2016), which are ideally capable of direct manipulation of the mediator, these are usually difficult to implement and are less informative than a traditional experiment, as the point estimates are not estimable in all cases. Pósch (2019) used parallel (encouragement) design in the context of procedural justice policing that exemplifies the challenges of these techniques.

Modelling multiple mediators at the same time is also a demanding exercise that requires either stricter parametric assumptions, more limited decomposition or complex model specification (Imai and Yamamoto 2013; Kim, Daniels, and Hogan 2018; VanderWeele and Vansteelandt 2014). Pósch (2019) published a review of causal mediation analysis techniques with multiple mediators. The motivating example focused on police legitimacy as mediators of the effect of procedural justice on willingness to cooperate with the police.

Causal interactions and effect heterogeneity

Studies with randomised experiments tend to focus on the 'first-generation question', which seeks to answer whether, and to what extent, a treatment had an effect on the outcome (i.e., average treatment effect). In comparison, a 'second-generation question' investigates whether there are contexts which accentuate or attenuate a treatment effect or whether there are certain individual characteristics which make people more or less affected by the treatment (Na, Loughran, and Paternoster 2015). In experimental studies, it is worth differentiating between three analytically equivalent but substantively different interactions: (a) design-induced causal interactions, (b) effect heterogeneity, and (c) causal interactions with third variables.

Factorial and balanced conjoint experimental designs provide the most straightforward cases of design-induced causal interactions (Aguinis and Bradley 2014; Egami and Imai 2016; Liebe et al. 2017). In such studies, multiple parameters are manipulated independently of each other and researchers are interested not only in the main effects of each parameter but also the potential causal dependency between them. A good example is Reisig, Mays, and Telep's (2018) experiment where procedural justice and the outcome of the encounter were independently manipulated. Because each parameter is randomly allocated to the respondents, the emerging interactions are guaranteed to have causal properties. Factorial designs can determine whether the effects of certain treatments are better combined for increased effect or they are independent and hence can be introduced separately.

Policy-makers often want to gain the best returns for their investment and effect heterogeneity can help identifying subpopulations which can benefit the most or least of a given intervention (Green and Kern 2012; Imai and Ratkovic 2013). For instance, evaluating the impact of the introduction of

parental leave on subjective well-being is expected to find a larger positive effect on women than men, who are more likely to stay home with their newborn(s) after giving birth. In such cases, the lack of heterogeneity would be a potential cause for alarm regarding the policy. It is also possible, however, that certain subpopulations can suffer as an unintended consequence of an otherwise beneficial treatment. Legewie and Fagan (2019) found that as a side-effect of 'Operation Impact', the increased police activity in New York's high crime areas had a statistically significant negative effect on educational performance, but only for older, male and African American students. Importantly, however, and unlike with factorial designs, effect heterogeneity cannot inform us whether the differential effects are due to the pre-treatment characteristic or another unmeasured factor associated with them. For instance, in case of Legewie and Fagan (2019) it is not possible to discern whether the change in effects was due to the students' gender, age, ethnicity, all of them combined, or whether they were only proxies for the more important variable(s) (VanderWeele and Knol 2014). Accordingly, knowing how the treatment effect varies across individuals or distinct groups can influence the decision which subpopulations (not) to target with a certain treatment for maximum returns. However, the emerging interactions (sometimes referred to as covariate average treatment effects or CATE) are not causal parameters. Another issue with treatment effect heterogeneity is that often a large number of covariates can be considered for interactions. Pósch's (2019) article uses the Scottish Community Engagement Trial and demonstrates how a machine learning algorithm can estimate treatment effect heterogeneity in randomised experiments.

A third and relatively rarely explored interpretation is causal interactions with third variables (Vanderweele 2009b, 2009a). This approach considers cases where the effect of the intervention might be dependent on a third variable that—despite not being randomised itself—possesses causal properties. Take for instance the paternal leave policy discussed earlier, but instead of using gender as the other variable, consider the number of hours spent resting. It is conceivable that the effect of the parental leave policy might be dependent on whether it managed to secure more hours of rest for the caretakers. If the effect of the treatment is only 'activated' in the presence of another third variable, these cases are referred to as 'sufficient cause interactions' or 'mechanistic interactions' (VanderWeele and Richardson 2012). Thus, the reason why someone might want to consider this interpretation of the interaction is that certain times only the interaction between a randomised treatment and a third variable can produce (an augmentation of) the desired outcome. Crucially, effect heterogeneity and causal interactions with third variables only differ in the assumptions used. The derived conditional average treatment effects can only be considered true causal estimates if the relationship between the parental leave policy, 'hours spent resting', and subjective well-being has no unmeasured confounding, which usually requires a long list of covariates being entered to the model. In addition, the monotonicity assumption needs to be satisfied which is also a strong and untestable assumption (i.e., the interaction between treatment and a third variable cannot have the opposite effect for any individual – e.g., the parental leave cannot reduce the time spent resting for anyone) (VanderWeele and Knol 2014).[2] Notably, and as with causal mediation analysis, this model can be extended to more than two-way interactions as well (Vanderweele 2009b).

A unified view of causal mechanisms and a demonstration

So far we have discussed causal mediation and causal interactions separately. When it comes to its relationship with perceived procedural fairness, perceived legitimacy can be either a mediator, a moderator, or both. As a mediator, prior levels of perceived procedural fairness of officers in general could be altered by direct or indirect experience with the police (specifically, the procedurally just or unjust actions of the officer involved), changing levels of general levels of perceived police procedural fairness, thereby changing levels of perceived legitimacy. In the next interaction with the police, perceptions of legitimacy could shape how people behave, thereby shaping the experience of procedural justice. As a moderator, legitimacy (specifically the deference part of the concept) may shape how people view the fairness of an officer in a direct or indirect encounter, thereby moderating the effect of that experience on subsequent levels of perceived procedural fairness in general (perhaps a more positive effect when deference is high and a less positive effect when deference is low).

[2] Although there have been models allowing to relax the monotonicity assumption, the standard tests of interaction do not apply to such cases (VanderWeele, 2009b).

As outlined by VanderWeele (2014), it is possible to test both hypotheses simultaneously by relying on a four-way decomposition of the average treatment effect. This approach divides the total effect by isolating four components, the sum of which will be equal to the average treatment effect. Notably, this decomposition is only viable provided that the sequential ignorability assumption (outlined in the causal mediation section) is satisfied. The four components are:

1. The controlled direct effect (CDE) or the unmediated and unmoderated part of the average treatment effect
2. The reference interaction (RI) or an additive interaction between the treatment and the intermediate variable
3. The pure indirect effect (PIE) or the main (mediated) effect that goes through the intermediate variable towards the outcome
4. The mediated interaction (MI) where the mediated effect is dependent on an interaction between the treatment and the intermediate variable

To demonstrate the utility of this approach, we use a recent online policing experiment where participants from the UK (Prolific Academic, n=438) were shown pictures of either plain-clothes civilians (control) or police officers in uniforms (treatment).[3] This experiment assessed whether the sheer (visual) presence of officers increased the legitimacy of the police (normative alignment and duty to obey). It had been assumed that any potential increase in legitimacy was due to stronger identification with the police (Bradford 2014; Bradford, Murphy, and Jackson 2014). In other words, social identification with the police was chosen as the intermediate variable to explain why and how the treatment affected police legitimacy.

Social identity was measured by a 'Strongly disagree-Strongly agree' 5-point Likert-scale, with three items: 'I identify with the police.', 'I feel similar to the police.', and 'I feel a sense of solidarity with the police.'. Normative alignment with the police was captured by a similar Likert-scale and three items: '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'. Finally, duty to obey the police was measured by a 5-point 'Not at all my duty-Completely my duty' Likert-scale. 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?". Component scores were derived for these three scales using principal component analysis. The following covariates were controlled for in all models: gender, age, ethnicity, foreign-born/UK-born, nationality (British/other), political orientation (left-right scale and attitude towards Brexit), and criminal justice experience (citizen- or police-initiated contact, and victimisation in the past two years). Linear modelling strategy was pursued, and the standard errors were estimated with 1,000 bootstrap samples. The results are shown in Table 1.

The average treatment effects of seeing pictures of police officers instead of civilians were significant both for normative alignment (ATE=0.340, $p<0.001$) and duty to obey (ATE=0.275, $p<0.001$) with a slightly higher effect size for the former. The decomposed ATEs of normative alignment and duty to obey showed a very similar picture. The unmediated and unmoderated part of the treatment effect (CDE) was significant for both normative alignment (CDE=0.152, $p<0.01$) and duty to obey (CDE=0.139, $p<0.01$). In comparison, the effect sizes of the reference interactions (normative alignment: RI=0.009, $p>0.05$; duty to obey: RI=-0.003, $p>0.05$) and mediated interactions (normative alignment: MI=-0.017, $p>0.05$; duty to obey: MI=0.006, $p>0.05$) were all non-significant and very close to zero. Simply put, there were no signs of causal dependencies either between the treatment and the intermediate variable, or the mediated effect, the treatment, and the intermediate variable. By contrast, the mediated effects on their own were significant for normative alignment (PIE=0.196, $p<0.001$), as well as duty to obey (PIE=0.133, $p<0.001$). These results imply, that the treatment effect of seeing pictures of police officers on police legitimacy was partially mediated by social identification with the

[3] The ethnicity of the officer was also manipulated but for the sake of simplicity we disregard this aspect now.

police. Nevertheless, the significant direct effects suggest that social identification only partially explains how and why seeing the pictures increased subjective police legitimacy.

Table 1: parameter estimates from the policing experiment

Effects	*Normative alignment*	*Duty to obey*
Average Treatment Effect (ATE)	0.340*** [0.224, 0.456]	0.275*** [0.171, 0.379]
Controlled Direct Effect (CDE)	0.152** [0.064, 0.240]	0.139** [0.050, 0.228]
Reference interaction (RI)	0.009 [-0.002, 0.019]	-0.003 [-0.013, 0.006]
Pure Indirect Effect (PIE)	0.196*** [0.112, 0.279]	0.133*** [0.077, 0.189]
Mediated Interaction (MI)	-0.017 [-0.035, 0.002]	0.006 [-0.012, 0.024]

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

Limitations: addressing 'third common causes' and 'reverse causality'

Due to their direct relevance, we think it is best to discuss the limitations of the proposed methods through revisiting Nagin and Telep's (2017) influential commentary on the state of the procedural justice literature. In their recent review, they argued that there is only limited and insufficient causal evidence for procedural justice policing. Despite Tyler's (2017) powerful response, we largely agree with Nagin and Telep's analysis, and we would encourage researchers to carry out more sophisticated causal analyses. Hence, we believe that it is useful to engage with two of Nagin and Telep's criticisms: the issue of 'third common causes' and 'reverse causality'.

Starting with 'third common causes', Nagin and Telep (2017: 7) voiced their scepticism whether police encounters can meaningfully affect subjective procedural justice and police legitimacy, or these attitudes cannot be disentangled from other influences, such as 'extreme poverty, racial isolation, and various forms of social dysfunction'. If indeed the perception of procedural justice and police legitimacy are determined by other factors, such as the neighbourhood context or legal socialisation irrespective of police action, this can be an insurmountable problem. However, well-executed RCTs can address this issue through random assignment which makes the treatment and control groups on average the same in all measured and unmeasured ways. In other words, random assignment can do away with third common causes making the average treatment effect a valid causal estimate. As pointed out in the discussion about causal interactions, this ought not to mean that certain personal characteristics (such as ethnicity) cannot have an influence creating potentially heterogeneous treatment effects. However, even in such cases, the available toolkit can readily examine and tackle the possibility of such effects (Green and Kern 2012; Imai and Ratkovic 2013).

Third, common causes are more problematic in the case of causal mediation analysis and causal interactions, where the relationship between the mediator/interaction and the outcome could be possibly caused by an unmeasured confounder. As this is an untestable assumption, it is impossible to ascertain whether this was the case. Yet, there are still ways to address this limitation, such as adding a long list of potentially influential covariates to the model (akin to 'comprehensive SEMs', MacKinnon and Pirlott 2015) and using sensitivity analysis techniques to quantify the robustness of the results (Cox et al. 2013; Imai et al. 2010; Pósch, 2019). Hence, future studies should establish sensitivity benchmarks making it possible to differentiate real and spurious effects.

The issue of 'reverse causality' – an oxymoron in the causal inference literature – only emerges in case of causal mediation analysis, where there is no empirical method which could tell whether the mediator caused the outcome or the other way around. There have been several unsuccessful attempts in the mediation analysis literature to quantitatively assess this problem (Spencer, Zanna, and Fong 2005; Wiedermann and Sebastian 2018) but most of them have been refuted (Bullock, Green, and Ha 2010; Lemmer and Gollwitzer 2017; Thoemmes 2015) with the main issue being that this is also a non-testable assumption. There have been notable developments in statistics that allow distinguishing between cause and effect (Mooij et al. 2016), but these rely on restrictive and unrealistic assumptions such as deterministic relationship (i.e., no measurement error) and no unmeasured confounding, which only makes them usable in the natural sciences such as astronomy.

An alternative way to test for 'reverse causality' is through establishing temporal order. Indeed, some argue that causal mediation analysis ideally needs to be longitudinal (Preacher 2015; Walters and Mandracchia 2017), however, others dispute the necessity of this (Imai et al. 2010; VanderWeele 2015). Yet longitudinal studies – even RCTs and traditional difference-in-differences analysis – are incapable of addressing the problem of 'reverse causality', because they can only control for time-invariant covariates such as age, gender, etc. (Imai and Kim 2019). There are ways of evaluating cases with time-variant treatment and mediation (Clare, Dobbins, and Mattick 2018; Daniel, De Stavola, and Cousens 2011), but these cannot solve the problem either. Importantly, statistical (non-)significance can only provide an indication of the presence/absence of a theorised relationship, but not clear evidence of it. This means that even experiments that were designed to manipulate the mediator, such as the parallel (encouragement) design (Imai 2013; Pósch, 2019), cannot rule out 'reverse causality'. Due to these intractable difficulties, and unless clear theoretical and temporal order can be irrefutably established, 'reverse causality' remains both an untestable and unsolvable problem. As a result, the best thing researchers can do is to get some indication (but not proof) from the statistical significance of the results and make their a priori theory-driven case based on the existing empirical evidence.

Discussion

We started this chapter by outlining some of the complexities of police-citizen encounters. We then considered the dominant theoretical account of police-citizen relations: namely, procedural justice theory (Sunshine & Tyler, 2003; Tyler, 2006a, 2006b; Tyler & Jackson, 2014). This led us to discuss two gaps in the literature. The first relates to fairness perception, particularly the idea (a) that different people can come to different conclusions regarding police fairness according to instrumental, relational and moral motives directing non-directional and directional goals, and (b) that social identity *and* personal sense of power may be two partial mediators of the procedural justice to legitimacy relationship.

In the third section we provided a brief roadmap for future analysis, and we demonstrated that causal mediation and causal interaction analyses are effective tools to test the model of procedural justice policing outlined earlier. In summary, there are at least five reasons to pursue this approach:

1. Causal mediation and causal interaction analyses originate in Structural Equation Modelling, which is a commonly used modelling technique in the procedural justice literature, making the concepts discussed familiar to people working on the field.
2. These methods utilise the potential outcome framework which provides rigour and clarity regarding the modelling and causal identifying assumptions.
3. The methods discussed here decompose the average treatment effects estimated for RCTs, thus directly building on and extending existing analytical strategies.
4. Sensitivity analysis techniques are available for most methods which can quantify the robustness of the effects in relation to certain causal identifying assumptions, such as 'unmeasured third causes'.
5. Finally, these models allow for more flexible modelling and weaker causal identifying assumptions compared to usual models of Structural Equation Modelling.

Nevertheless, we want to end our chapter with a word of caution regarding causal mechanisms. Due to the stringent no unmeasured confounding assumptions, we do not recommend assessing causally mediating and moderating effects as an exploratory exercise. As with all rigorous scientific research, the first steps ought to be theory building, followed by empirical tests of associations, then assessment of the average treatment effects, and, only then should causal mechanisms be sought. Without extensive knowledge and clear understanding of the place of a construct in a broader causal hierarchy, it is difficult to determine whether emerging effects have true causal properties or are caused by influential unmeasured third sources. Therefore, scrutinising causal mechanisms is only advisable in case of well-developed theories that have been severely examined and are supported by ample empirical evidence. Fortunately, procedural justice policing is one of these mature theories, where testing causal mechanisms is a natural next step.

References

Aguinis, H. and K. J. Bradley. 2014. "Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies." Organizational Research Methods 1094428114547952-.

Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. Psychological Bulletin, 141(3), 574.

Barclay, L. J., Bashshur, M. R., & Fortin, M. (2017). Motivated cognition and fairness: Insights, integration, and creating a path forward. Journal of Applied Psychology, 102(6), 867-889.

Barkworth, J. M., & Murphy, K. (2015). Procedural justice policing and citizen compliance behaviour: The importance of emotion. Psychology, Crime & Law, 21(3), 254-273.

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.

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

Bittner, E. and Bish, R. L. (1975). The functions of the police in modern society : a review of background factors, current practices, and possible role models, Seattle, University of Washington Press.

Bittner, E. (1967). The Police on Skid-Row: A Study of Peace Keeping. American Sociological Review, 32, 699-715.

Bittner, E. (2005). Florence Nightingale in pursuit of Willie Hutton. In: NEWBURN, T. (ed.) Policing : key readings. Cullompton, Devon ; Portland, Or.: Willan Pub.

Blader, S. L., & Tyler, T. R. (2003). A four-component model of procedural justice: Defining the meaning of a "fair" process. Personality and Social Psychology Bulletin, 29(6), 747-758.

Blader, S. L., & Tyler, T. R. (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), 445-464.

Blader, S. L., & Tyler, T. R. (2015). Relational models of procedural justice. The Oxford handbook of justice in the workplace, 351, 370.

Bradford B, Milani J, Jackson J. 2017. Identity, legitimacy and "making sense" of police use of force. Polic. Int. J. Police Strateg. Manag. 40(3):614–27

Bradford, B., Jackson, J. and Hough, M. (2013). 'Police Futures and Legitimacy: Redefining "Good Policing"', in Brown, J. (ed.) *The Future of Policing.* Oxon: Routledge, pp. 79-99. * Invited position paper for the Stevens' Independent Commission on the Future of Policing in England & Wales.

Bradford, B., Jackson, J. and Hough, M. (2014). 'Police Legitimacy in Action: Lessons for Theory and Practice', in Reisig, M. & Kane, R. (eds.) *The Oxford Handbook of Police and Policing*. Oxford: Oxford University Press, pp. 551-570.

Bradford, B., Jackson, J. and Stanko, E. (2009). 'Contact and Confidence: Revisiting the Impact of Public Encounters with the Police', Policing and Society, 19, 1, 20-46,

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

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

Brockner, J., & Greenberg, J. (1990). The impact of layoffs on survivors: An organizational justice perspective. Applied social psychology and organizational settings, 45, 75.

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.

Carson, Jennifer V. and Ashley P. Wellman. 2017. "Problem-Oriented Policing in Suburban Low-Income Housing: A Quasi-Experiment." Police Quarterly 109861111774400.

Chaikin, A. L., & Darley, J. M. (1973). Victim or perpetrator?: Defensive attribution of responsibility and the need for order and justice. Journal of Personality and Social Psychology, 25(2), 268.

Clare, Philip J., Timothy A. Dobbins, and Richard P. Mattick. 2018. "Causal Models Adjusting for Time-Varying Confounding — a Systematic Review of the Literature." International Journal of Epidemiology 1–12.

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, 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.
de Puiseau, B. W., Glöckner, A., & Towfigh, E. V. (2019). Integrating theories of law obedience: How utility-theoretic factors, legitimacy, and lack of self-control influence decisions to commit low-level crimes. *Judgment & Decision Making*, *14*(3).
Egami, Naoki and Kosuke Imai. 2016. "Causal Interaction in Factorial Experiments: Application to Conjoint Analysis." ArXiv.
Fairfield, Tasha and Andrew Charman. 2017. "Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats." Political Analysis 25(3):363–80.
Geller, A., & Fagan, J. (2019). Police contact and the legal socialization of urban teens. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *5*(1), 26-49.
Goffman, E. 1958. The presentation of self in everyday life, Edinburgh, University of Edinburgh, Social Sciences Research Centre.
Goffman, E. 1961. Encounters : two studies in the sociology of interaction, Indianapolis, Bobbs-Merrill.
Goffman, E. 1967. Interaction ritual : essays in face-to-face behavior, London ; Chicago, The Penguin Press : Aldine.
Green, Donald P. and Holger L. Kern. 2012. "Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees." Public Opinion Quarterly 76(3):491–511.
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.
Hassell, Kimberly D. and Rickie D. Lovell. 2015. "Fidelity of Implementation: Important Considerations for Policing Scholars." Policing and Society 25(5):504–20.
Hedström, Peter and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." Annual Review of Sociology 36(1):49–67.
Hough, M., Jackson, J., Bradford, B ., Myhill, A. and Quinton, P. (2010). 'Procedural Justice, Trust and Institutional Legitimacy', *Policing: A Journal of Policy and Practice*, 4, 3, 203-210.
Hough, M., Jackson, J. and Bradford, B. (2013). 'The Governance of Criminal Justice, Legitimacy and Trust', in Body-Gendrot, S., Lévy, R., Hough, M. Snacken, S. and Kerezsi, K. (eds.) *The Routledge Handbook of European Criminology*. Oxon: Routledge, pp. 243-265.
Huq, A., Jackson, J. and Trinkner, R. (2017). 'Legitimating Practices: Revisiting the Predicates of Police Legitimacy', British Journal of Criminology, 57, 5, 1101-1122.
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 In Song Kim. 2019. "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" American Journal of Political Science 00(0):1–24.
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.
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. 2013. "Experimental Designs for Identifying Causal." 5–51.
Jackson, J. (2018). 'Norms, Normativity and the Legitimacy of Legal Authorities: International Perspectives', Annual Review of Law and Social Science, 14, 145-165.

Jackson, J., Bradford, B., Hough, M., Myhill, A., Quinton, P., and Tyler, T. R. (2012). 'Why do People Comply with the Law? Legitimacy and the Influence of Legal Institutions', British Journal of Criminology, 52, 6, 1051-1071,

Jackson, J., Bradford, B., Stanko, E. A. and Hohl, K. (2013). Just Authority? Trust in the Police in England and Wales. Oxon: Routledge.

Jackson, J., Huq, A., Bradford, B. and Tyler, T. R. (2013). 'Monopolizing Force? Police Legitimacy and Public Attitudes towards the Acceptability of Violence', Psychology, Public Policy and Law, 19, 4, 479-497.

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.

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

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.

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.

Legewie, Joscha and Jeffrey Fagan. 2019. "Aggressive Policing and the Educational Performance of Minority Youth." American Sociological Review 000312241982602.

Lemmer, Gunnar and Mario Gollwitzer. 2017. "The 'True' Indirect Effect Won't (Always) Stand up: When and Why Reverse Mediation Testing Fails." Journal of Experimental Social Psychology 69:144–49.

Liebe, Ulf, Ismaïl M. Moumouni, Christine Bigler, Chantal Ingabire, and Sabin Bieri. 2017. "Using Factorial Survey Experiments to Measure Attitudes, Social Norms, and Fairness Concerns in Developing Countries." Sociological Methods & Research.

Lind, E. A. (2001). Fairness heuristic theory: Justice judgments as pivotal cognitions in organizational relations. In J. Greenberg and R. Cropanzano (Eds.), *Advances in organizational justice* (pp. 56-88). Palo Alto, CA: Stanford University Press.

Lind, E. A., & Tyler, T. R. (1988). The social psychology of procedural justice. Springer Science & Business Media.

Lind, E. A., Kray, L., & Thompson, L. (2001). Primacy effects in justice judgments: Testing predictions from fairness heuristic theory. Organizational behavior and human decision processes, 85(2), 189-210.

Lowrey, B. V., Maguire, E. R., & Bennett, R. R. (2016). Testing the effects of procedural justice and overaccommodation in traffic stops: A randomized experiment. Criminal justice and behavior, 43(10), 1430-1449.

MacKinnon, D. P. and A. 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. 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.

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

Mazerolle L, Antrobus E, Bennett S, Tyler TR. 2013. Shaping citizen perceptions of police legitimacy: a randomized field trial of procedural justice. Criminology 51:33–63

Mooij, J. M., J. Peters, D. Janzing, J. Zscheischler, and B. Schölkopf. 2016. "Distinguishing Cause from Effect Using Observational Data: Methods and Benchmarks." Journal of Machine Learning Research 17:1–102.

Murphy, T., Bradford, B. and Jackson, J. (2016). 'Motivating Compliance Behavior Among Offenders: Procedural Justice or Deterrence?', *Criminal Justice & Behavior*, 43, 1, 102-118

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 DS, Telep CW. 2017a. Procedural justice and legal compliance. Annu. Rev. Law Soc. Sci. 13:5–28

Nagin DS, Telep CW. 2017b. Reponse to "Procedural Justice and Policing: A Rush to Judgment?" Annu. Rev. Law Soc. Sci. 13:55–58

Ogburn, Elizabeth L., Oleg Sofrygin, Ivan Diaz, and Mark J. van der Laan. 2017. "Causal Inference for Social Network Data." Unpublished Manuscript 1–43.

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.

Papachristos, V. Andrew, Jeffrey Fagan, Meares, L. Tracey, Fagan, and Jeffrey. 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.

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.

Posch, K., Jackson, J., Bradford, B. and MacQueen, S. (2019). '"Truly Free Consent"? Clarifying the Nature of Police Legitimacy using Causal Mediation Analysis', SocArXiv.

Pósch, Krisztián. 2019. "Testing Complex Social Theories With Causal Mediation Analysis and G-Computation: Toward a Better Way to Do Causal Structural Equation Modeling." Sociological Methods and Research.

Preacher, Kristopher J. 2015. "Advances in Mediation Analysis: A Survey and Synthesis of New Developments." Annual Review of Psychology 66:825–52.

President's Task Force on 21st Century Policing, 2015).

Radburn, M., & Stott, C. (2018). The social psychological processes of 'procedural justice': Concepts, critiques and opportunities. Criminology & Criminal Justice, 1748895818780200.

Radburn, M., Stott, C., Bradford, B., & Robinson, M. (2018). When is policing fair? Groups, identity and judgements of the procedural justice of coercive crowd policing. Policing and Society, 28(6), 647-664.

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.

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.

Skarlicki, D. P., O'Reilly, J., & Kulik, C. T. (2015). The third-party perspective of (in)justice. In R. Cropanzano & M. L. Ambrose (Eds.), Oxford handbook of justice in the workplace (pp. 235–255). Oxford, England: Oxford University Press.

Solomon, S. J. (2019). How Do the Components of Procedural Justice and Driver Race Influence Encounter-Specific Perceptions of Police Legitimacy During Traffic Stops?. *Criminal Justice and Behavior*.

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.

Stanko, E. A., Jackson, J., Bradford, B. and Hohl, K. (2012). 'A Golden Thread, a Presence amongst Uniforms, and a Good Deal of Data: Discourses of Confidence in the London Metropolitan Police', *Policing and Society*, 22, 3, 317-331.

Sunshine J, Tyler TR. 2003. The role of procedural justice and legitimacy in public support for policing. Law Soc. Rev. 37:513–48

Thibaut, J. W., & Walker, L. (1975). Procedural justice: A psychological analysis. Hillsdale, NJ: Erlbaum.

Thoemmes, Felix. 2015. "Reversing Arrows in Mediation Models Does Not Distinguish Plausible Models." Basic and Applied Social Psychology 37(4):226–34.

Trinkner, R., Jackson, J. and Tyler, T. R. (2018). 'Bounded Authority: Expanding "Appropriate" Police Behavior Beyond Procedural Justice', Law & Human Behavior, 42, 3, 280-293
Trinkner, R., Mays, R. D., Cohn, E. S., Van Gundy, K. T., & Rebellon, C. J. (2019). Turning the corner on procedural justice theory: exploring reverse causality with an experimental vignette in a longitudinal survey. Journal of Experimental Criminology, 1-11.
Tyler TR, Fagan J. 2008. Why do people cooperate with the police? Ohio State J. Crim. Law 6:231–75
Tyler TR. 2006a. Why People Obey the Law. New Haven, CT: Yale Univ. Press. 2nd ed.
Tyler TR. 2006b. Psychological perspectives on legitimacy and legitimation. Annu. Rev. Psychol. 57:375–400
Tyler, T. R. (2009). Legitimacy and criminal justice: The benefits of self-regulation. Ohio St. J. Crim. L., 7, 307.
Tyler, T. R., & Blader, S. L. (2001). Identity and cooperative behavior in groups. Group Processes & Intergroup Relations, 4(3), 207-226.
Tyler, T. R., & Blader, S. L. (2003). The group engagement model: Procedural justice, social identity, and cooperative behavior. Personality and social psychology review, 7(4), 349-361.
Tyler, T. R., & Lind, E. A. (1992). A relational model of authority in groups. In Advances in experimental social psychology (Vol. 25, pp. 115-191). Academic Press.
Tyler, T. R., Goff, P. A., & MacCoun, R. J. (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 R. 2017. "Procedural Justice and Policing: A Rush to Judgment?" Annual Review of Law and Social Science 13:29–53.
Tyler, T. R. (2017). Can the police enhance their popular legitimacy through their conduct: using empirical research to inform law. *U. Ill. L. Rev.*, 1971.
Van den Bos, K., Lind, E. A., and Wilke, H. (2001). The psychology of procedural justice and distributive justice viewed from the perspective of fairness heuristic theory. In R. Cropanzano (Ed.), *Justice in the workplace: Volume II—From theory to practice* (pp. 49-66). Mahwah, NJ: Lawrence Erlbaum & Associates.
Vanderweele, Tyler J. 2009a. "On the Distinction between Interaction and Effect Modification." Epidemiology 20(6):863–71.
Vanderweele, Tyler J. 2009b. "Sufficient Cause Interactions and Statistical Interactions." Epidemiology 20(1):6–13.
Vanderweele, Tyler J. 2014. "A Unification of Mediation and Interaction: A Four-Way Decomposition." Epidemiology 25(5):749–61.
VanderWeele, Tyler J. 2015. Explanation in Causal Inference - Methods for Mediation and Interaction. Oxford University Press.
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.
VanderWeele, Tyler J. and Thomas S. Richardson. 2012. "General Theory for Interactions in Sufficient Cause Models with Dichotomous Exposures." The Annals of Statistics 40(4):2128–61.
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.
Waddington, P. A. J. 1999a. Discretion, 'Respectability' and Institutional Police Racism. Sociological Research Online, 4.
Waddington, P. A. J. 1999b. Policing citizens : authority and rights, London, UCL Press.
Waddington, P. A., Wright, M., Williams, K., & Newburn, T. (2018). How people judge policing. Oxford University Press.
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.
Weller, Nicholas and Jeb Barnes. 2016. "Pathway Analysis and the Search for Causal Mechanisms." Sociological Methods & Research 45(3):424–57.

Wiedermann, Wolfgang and James Sebastian. 2018. "Direction Dependence Analysis in the Presence of Confounders : Applications to Linear Mediation Models Using Observational Data." Multivariate Behavioral Research 0(0):1–20.

Weisburd, D., Majmundar, M. K., Aden, H., Braga, A., Bueermann, J., Cook, P. J., ... & Manski, C. (2019). Proactive Policing: a Summary of the Report of the National Academies of Sciences, Engineering, and Medicine. *Asian Journal of Criminology*, *14*(2), 145-177.