A simple checklist, that is all it takes: a cluster randomized controlled field trial on improving the treatment of suspected terrorists by the police

Brandon Langley[1] · Barak Ariel[1] · Justice Tankebe[1] · Alex Sutherland[3] · Marcus Beale[4] · Roni Factor[2] · Cristobal Weinborn[5]

Abstract

Objectives When it comes to interviewing suspected terrorists, global evidence points to harsh interrogation procedures, despite the likelihood of false positives. How can the state maintain an effective counterterrorism policy while simultaneously protecting civil rights? Until now, the shroud of secrecy of "national security" practices has thwarted attempts by researchers to test apparatuses that engender fair interrogation procedures. The present study aims to test one approach: the use of a "procedural justice checklist" (PJ Checklist) in interviews of suspected terrorists by counterterrorism police officers in port settings.

Methods Using a clustered randomized controlled field test in a European democracy, we measure the effect of implementing Procedural Justice (PJ) Checklists in counter-terrorism police settings. With 65 teams of officers randomly-assigned into treatment and control conditions, we compare post-interrogation surveys of suspects ($n = 1418$) on perceptions of legitimacy; obligations to obey the law; willingness to cooperate with the police; effectiveness of counterterrorism measures; distributive justice; feelings of social resistance to the state; and PJ. A series of multi-level linear, logistic, and ordered logit regression models are used to estimate the treatment effect, with Hedges' g and odds ratios used for effect sizes.

Results When compared with control conditions, implementing a policy of PJ Checklist causes statistically significant and large enhancement in all measured dimensions, including the willingness of suspects to obey the law ($g = 1.022$ [0.905, 1.138]), to cooperate with the police ($g = 1.118$ [0.999, 1.238]), distributive justice ($g = 0.993$ [0.880, 1.106]), effectiveness ($g = 1.077$ [0.959, 1.195]), procedural justice ($g = 1.044$ [0.930, 1.158]), and feelings of resistance towards the state ($g = -0.370$ [-0.259, -0.482]).

✉ Barak Ariel
ba285@cam.ac.uk; barak.ariel@mail.huji.ac.il

Extended author information available on the last page of the article

🌀 Springer

Conclusions PJ checklists offer a simple, scalable means of improving how state agents interact with terrorism suspects. The police can use what is evidently a cost-effective tool to enhance legitimacy and cooperation with the police, even in a counterterrorism environment.

If asked "How should society treat terrorists?", most people would argue for a more punitive response (Hood and Hoyle 2015). Some citizens are not upset when security forces do "whatever it takes" to protect them from the threat of terrorism, even if using "extreme measures" implies or creates violations of human rights (Kugler and Cooper 2010). In these scenarios, political violence and the defense of civil liberties come into conflict, leading to serious ethical and psychological discord (Merin et al. 2015).

Every human being deserves to be treated with dignity and respect, but this seems especially to be the case when the culpability of the suspect has yet to be demonstrated (Europol 2017; Newheiser and DeMarco 2018). After all, suspects might simply be innocent civilians, and the initial suspicion may have been erroneous. Treating people who did not commit any crime unfairly is a major concern, as doing so can lead to reduced belief in the legitimacy of the law, by the general public, those suspected of terrorism, or both (LaFree 2018). It may also alienate the group to which the suspected terrorist belongs, which could backfire when asking for help to counter terrorism—the most troublesome manifestation of this being that some may be driven to extremism as a reaction to the harsh counterterrorism approach against their community (Cronin 2006).

Despite these concerns, there is global evidence that suspected terrorists are likely to be treated unfairly in the name of national security interests (Guiora 2006; Piazza and Walsh 2010). Wherever we find actual or perceived terrorism threats, we see evidence of repressive strategies (Dreher et al. 2010). Common features of counterterrorism policy include "aggressive interrogation techniques," procedural impropriety, and neglect of fundamental rights and liberties (Lindekilde and Sedgwick 2012), even in societies with a longstanding tradition of protecting human rights (Bayley and Weisburd 2009; Steyn 2004).

What can be done to improve the ways in which state actors interact with suspected terrorists? At the very least, how can we improve the conduct of security enforcement agencies when they interview suspected terrorists? This is not an easy task, as much of the work of these agencies remains "in the shadows," often out of reach to researchers. The public usually learns about supposed adverse interrogation room tactics and the removal of due process rights on an ad hoc basis (Shane 2009). Likewise, researchers have found it challenging to access the world of counterterrorism; as a result, the amount of evidence-based interventions is limited (Lum et al. 2006). Consequently, we encounter a paradox: public safety situations in which greater certainty about a strategy's effectiveness could prove beneficial are precisely the situations in which evidence is scarce. As such, social scientists are rarely in a position to make evidence-based policy recommendations that might shape how security services respond to individuals suspected of terrorism (Ajzenstadt and Ariel 2008).

In this study, we present the results of the first national test of one type of intervention: a prospective experiment in which counterterrorism officers were tasked with applying a procedural justice checklist during their engagement with suspected terrorists. This checklist is similar to those that have been applied by pilots, surgeons, and NASA astronauts (see Gawande 2009). This reminder to comply with a rigid "code of practice" meant providing suspects with information about their arrest, allowing them to voice their concerns, and treating them impartially. The checklist was hypothesized to alter how suspects perceive interactions with counterterrorism officers across multiple dimensions, as well as their overall approval of policing.

What do we know about the treatment of suspected terrorists?

Research evidence shows that most people experience "morality silence" when asked how to respond to individuals suspected of terrorism (Kugler and Cooper 2010), i.e., a tendency to look the other way when the procedural rights of "enemies" are denied. These rights—ranging from freedom from torture, to having reasonable access to legal aid through pre-trial arrests or preventative detention, to respectful treatment by law enforcement officers—are commonly subverted worldwide (Kreps 2014). This moral and procedural ambivalence is most stark when it comes to the treatment of suspected and convicted terrorists (Gronke et al. 2010; Hewitt 1990). The greater people's sense of insecurity, the lower their support for civil liberties (Davis and Silver 2004; Viscusi and Zeckhauser 2003). Similarly, the indiscriminate violent acts perpetuated by terrorists severely provoke us, often leading to greater support for a "tougher hand" (Randahl 2018).

Given the above, it is unsurprising that police and security services use the full extent of legal measures to identify terrorists or gather intelligence about future terrorist attacks (Bamford 2005; Darmer 2002; Godsey 2005; Gross 2001). Further, in the context of an imminent attack, officials sometimes consider it justifiable to use more extreme approaches (Brecher and Brecher 2007; Dershowitz 2002). However, it becomes challenging to justify unusual techniques for the purpose of *identifying* a suspected terrorist. There are three objections commonly made to doing so, which we set out below.

First, applying tough or extreme measures is difficult to justify when the subject is merely a suspect or (ultimately) innocent (Curzer 2006). Law enforcement officers, like members of any profession, are fallible, and both clinical and actuarial assessments are prone to false positives and false negatives (Jonas and Harper 2006; Spinney 2010; Wainer and Savage 2008). Poor intelligence can lead to inaccurate identification of alleged terrorists, sometimes in large numbers (Chang 2011; Gross 2004) and can produce vast pools of "persons of interest" that present significant challenges for both police and security services. Law enforcement agencies should therefore apply clear codes of practice with regard to the fair treatment of suspects (Dotan 2004; Perito and Parvez 2014). A strong supporting argument for this practice is utilitarian: the more often that extreme measures are applied on false positives (i.e., arresting many but uninvolved members of the community), the less likely that community will be to support or assist with counterterrorism efforts (Hasisi et al. 2009).

Second, obtaining information under duress or via manipulation or illegitimate means often leads to false information (Malloy et al. 2014; Sangero 2016). Following

up on such leads means that actual terrorists remain at large and can continue carrying out attacks (Ginbar 2008). It is entirely conceivable that suspects will provide false information about themselves, or implicate the wrong persons, just "to give up some names", only as a way to stop the aggressive interrogation from continuing. Such false confessions and erroneous intelligence can hinder the investigation.

Third, there is a fundamental issue here: suspects should be treated professionally, respectfully, and without prejudice or bias (Grant 2007; Hirose 2014; Tankebe 2013; Watson 2019). Even if we assume that punishment for terrorism should be severe, the central rule in most democracy-based legal systems is that suspects should be treated fairly, no matter how heinous the crime of which they are suspected (McCrudden 2008). Until proven otherwise, the underlying assumption of law enforcement agencies must be that the suspect is innocent (Dotan 1999). We expect police and security services to handle such persons with respect and to accord them fundamental human rights (Rothe and Muzzatti 2004; Tyler and Wakslak 2004).

However, despite agreement about the need to treat suspects and the accused with respect (Obama 2010; Office of the United Nations High Commissioner for Human Rights 2008; Tyler 2012), the practice can fall far short of this aim. Evidence collected around the globe suggests that suspected terrorists are often treated disrespectfully or worse (Bonner 2013; Amnesty International 2017; Leslie 2017). Different ethnic, religious, or racial groups report in surveys that they experience discrimination, bias, and unfair treatment and are viewed as a security threat (Perry and Jonathan-Zamir 2014). Interestingly, the same groups of respondents often express strong and positive attitudes towards *cooperating* with the police and report terrorist threats to authorities (Hasisi and Weisburd 2011).

What can be done to minimize the risk of improper conduct during initial interviews?

Scholars interested in changing these practices highlight the need for training (Williamson 2013), better recruitment processes (Perliger et al. 2009), increased legal oversight (Guiora 2005), videotaping interrogations (Drizin and Colgan 2000; Hyatt et al. 2017), and conducting public inquiries (Rehman 2007). However, there is a limited number of rigorous evaluations of counterterrorism strategies, and virtually none concerning matters that take place during interviews. A systematic review of counterterrorism interventions detected only seven studies that used at least moderately rigorous scientific methods (Lum et al. 2006). The shroud of secrecy that characterizes counterterrorism strategies limits access to field settings and data, let alone carefully controlled tests of tactics (Langley 2014). We know of no randomized controlled trials that have attempted to modify the behavior of interviewers and detectives as they interact with suspected terrorists.

One recent approach that has been suggested as a tool for changing how policing practice is conducted is "procedural justice" (Tyler 2006). This term refers to perceptions of the fairness of processes that police officers—or, more broadly, legal authorities—use when they interact with citizens (Jonathan-Zamir and Harpaz 2018). Surveys in various countries have shown that, in forming their views about the legitimacy of police authority, citizens sometimes focus more heavily on the *quality of their interactions* with the police and much less on the *outcomes received* from the

police (Sunshine and Tyler 2003; Tyler 2006; Tyler et al. 2010). If citizens feel as though they were treated with dignity and respect and if they perceive the officer's decision as fair, they will trust the officer's decision (Tyler and Bies 1990) and feel a greater obligation to obey the law (Tyler and Wakslak 2004). Furthermore, research on police–citizen contacts involving the use of procedural justice shows a link to enhanced quality of police–citizen interactions, with citizens reporting greater satisfaction with the interaction (Mastrofski et al. 1996; McCluskey 2003; Reiss 1971; Reiss and Farrington 1991; Tyler and Fagan 2008), willingness to comply with police directives (Tyler and Bies 1990), and willingness to cooperate with the police (Murphy et al. 2008; Sunshine and Tyler 2003; Tyler and Wakslak 2004; Tankebe 2013).

Procedural justice can be contextualized more broadly within legitimacy theory. Legitimacy in law enforcement refers to the moral right of the state and its actors to govern. As such, it is multidimensional (Bottoms and Tankebe 2012). In the context of counterterrorism interviews, this intricacy is even more pronounced. If the police investigation should be viewed as "legitimate," for whichever purpose, then we would require the police officer to behave lawfully, to apply fair procedures when dealing with people, and to produce outcomes that are equally distributed between individuals of different backgrounds, but also to be effective in its mission (see review in Ariel et al. 2020).

A crucial gap in the literature is methodological: The majority of evidence on procedural justice and legitimacy is observational in design (Murphy et al. 2008; Tyler et al. 2014). A small number of studies applied randomized controlled trial methodologies to assess the effects of police behavior on legitimacy outcomes. These include the Queensland Community Engagement Trial in Australia (Mazerolle et al. 2013a) and the Scotland Community Engagement Trial (MacQueen and Bradford 2015), which involved use of "procedural justice scripts" by officers conducting breath-testing at roadblocks. The Australian experiment indicated that the procedurally just traffic stops improved the citizens' perceptions of both the individual officer and the entire police department (Mazerolle et al., 2013). Notably, however, in the Scotland study, citizens' perceptions of police legitimacy diminished with procedurally just police contact (MacQueen and Bradford 2015; but cf. MacQueen and Bradford 2017).

Similarly, Ariel et al. (2020) and Demir et al. (2018) used body-worn cameras to increase the application of procedural justice by traffic officers in Uruguay and Turkey, respectively. In the Uruguay study (Ariel et al. 2020; see also Mitchell et al. 2018), drivers who interacted with officers using body-worn cameras were more likely to report that they were listened to before the officers made decisions about their case, were treated with respect, and felt that officers expressed sincere concern for their well-being. They were also more likely to say that the traffic officers demonstrated neutrality in handling their cases. Collectively, the studies indicated significant improvements across various dimensions of legitimacy when officers using body-worn cameras issued tickets to traffic violators; importantly, these enhancements were linked to increases in perceived procedural justice, compared to control conditions (see Ariel et al. 2020).

How can procedural justice be increased during police interviews?

Making abstract concepts like "fairness," "legitimacy," "respect," and "fair treatment" actionable and measurable has been challenging. We know that samples of Muslims in

New York City (Tyler et al. 2010) and London (Huq et al. 2011), Arab Israelis in Israel (Hasisi and Weisburd 2011; Perry and Hasisi 2018), and passengers in U.S. airports (Lum et al. 2015; Sindhav et al. 2006) are concerned about procedural justice when it comes to counterterrorism. However, how to operationalize these concepts is unclear. The change is not necessarily legal, because in most democracies the lawfulness of counterterrorism is not as much a concern as its practice—i.e., black letter laws often prohibit disrespectful and unfair treatment of citizens (Clayton et al. 2000). In fact, a review of standard operating guidelines in law enforcement agencies worldwide shows that the manuals are, prima facie, within the rule of law—i.e., fair and unbiased—and the issue has always been in the application of these concepts (Schmid 2013).

Furthermore, the evidence regarding what works to increase the use of procedural justice in policing is limited and often mixed. Research tends to show an initial effect of police training, but this effect decays over time (Antrobus et al. 2018; Murphy et al. 2014). Enhanced oversight to increase the use of procedural justice is promising (Hyatt et al. 2017), though not without a potentially negative effect on the motivation and productivity of police officers (Wallace et al. 2018; Ariel et al. 2020).

Borrowing from the well-known work of Gawande (2009), we propose a different approach: using a checklist to remind officers of their code of practice during interviews with the intention of improving procedural compliance. Trials in various professions that implemented a checklist policy have shown an increase in compliance with rules and procedures on critical issues (Hales and Pronovost 2006; Weiser et al. 2010). Pilots (Degani and Wiener 1993), surgeons (Treadwell et al. 2014), astronauts (Marshburn et al. 2003), and physicians (Huis et al. 2012) were asked to use checklists to increase cooperation with specialized yet critical guidelines, and the results were similar: enhanced services and professional delivery of outcomes.

If checklists are useful in promoting health and safety, could they also be used to remind officers to be courteous, fair, and unbiased towards citizens in high-stakes interviews? Some evidence suggests that different forms of reminders are effective compared with no-reminder conditions (Langley 2014; Mazerolle et al. 2013a; Mitchell et al. 2018; Sahin et al. 2017). However, whether a checklist will successfully increase the fair and just treatment of suspected terrorists is presently unknown.

Methods

Field experiments in policing are still relatively rare, and controlled tests on counterterrorism are even rarer (Lum et al. 2006). We found no published trials applying more just or fair counterterrorism or counter-extremism processes. We were particularly interested in studies of counterterrorism powers that allow the police to operate with broad and intrusive powers to stop, search, and hold individuals suspected of being terrorists, as these powers are at the heart of potential human rights violations (Heath-Kelly 2012; Moran 2005). These laws provide security services with the discretion to stop, interview, or investigate individuals suspected of involvement in terrorism without the need for the same level of grounds required for "ordinary crimes" (see more broadly Ajzenstadt and Ariel 2008). However, we found no record of field evaluation with randomized controlled tests of these policies and practices.

Despite the lack of rigorous evaluations, in many countries, the use of such special policing powers is high at points of entry into a country, whether by sea, air, or land (hereinafter, "ports"). Counterterrorism officers are given the authority to detain and question any person, search them and their belongings, and take biometric information from them, usually without having to ask for permission. Further, it is an offense to remain silent in these interviews, and those being detained at the port have no right to a lawyer. These special powers apply when counterterrorism officers search for specific individuals based on certain intelligence upon the request of the security services (local or abroad). For example, the intelligence community might be aware of a suspected terrorist traveling between countries and alert the counterterrorism officers that a subject of interest is arriving. This has been the case with ISIS fighters who spent time in Syria and Iraq and are now returning to their home countries (Byman 2015). More broadly, the fact that some terrorists travel intercontinentally suggests that spatial clustering is modus operandi-specific (see Hasisi et al. 2019), indicating a wider range of terrorism threats and responses.

These powers to detain, intercept, or arrest can also be applied based on actuarial predictions and behavioral detection tactics, although these are likely to produce false positives (Mueller and Stewart 2016; Stewart and Mueller 2013). This is because identifying a person as a suspect at a port based on body language or overall demeanor and appearance is very difficult (Rae 2012). The likelihood of falsely identifying an innocent person as a suspected terrorist is thus very high in such settings. Take the United Kingdom as an example: The special powers to detain and question passengers in ports are found in Schedule 7 of the England and Wales Terrorism Act 2000 ('Schedule 7'), which is a national security port and border power that enables examining officers to stop, search, question and detain any individual traveling through a port. According to published information by the Home Office, 69,109 people were stopped and examined in 2011 and 2012 under Schedule 7 powers, of whom 681 were then detained. If one takes detention as a mark of success, the true positive rate is 0.0099% (Office for Security and Counter-Terrorism 2012). Given the seemingly small chance of apprehending a suspected terrorist without prior intelligence, and the likely high false positive rate, it seems that U.K. counterterrorism officers ought to treat people with whom they interact courteously, fairly, and respectfully, as they are likely to be innocent passengers rather than terrorists.

Study design

In the areas of counterterrorism, anti-extremism, counterinsurgency, and civil unrest, there are no rigorous evaluations using random allocation of treatments; see the debate over the use of these designs in counterterrorism in Laycock (2012) and Lum and Kennedy (2012), as well as Lum et al. (2006, p. 510–512). Other designs in this area are asymptotically able to infer causality like randomized controlled trials—see, e.g., agent-based modeling (Andrighetto et al. 2019). However, the present study conditions lent themselves to a controlled test of the effect of procedural justice on various outcomes associated with legitimacy. As detailed herein, the random allocation of units was both feasible and ethical, and the assumptions of randomized experiments were met—for example, having a sufficiently large sample size, controlling for self-selection

biases, maintaining treatment integrity, avoiding crossover contamination, and the availability of valid and reliable measurable outcomes (see review in West 2009).

This study was designed as a cluster-randomized controlled field trial to estimate the effect of a procedural justice checklist (PJ Checklist) on multiple outcomes compared with controlled conditions. We collaborated with 21 ports in this European democracy in which 65 teams were the units of randomization (hence the cluster component). We applied PJ Checklists in about half of the teams, while control teams proceeded with "business as usual." There was no spillover of cases, officers, or spatiotemporal interactions between the treatment and control teams. All officers in these clusters took part in the experiment (either as control or treatment), with no exclusions.

Settings

The experiment was initiated by the research team, who contacted the country's relevant counterterrorism authorities with a request to conduct the study. The 21 ports in this European democracy participated in a 12-month experiment beginning in March 2016, following a request sent to all sea, air, or land ports across the nation (over 100 in total). The approach to the ports with an invitation to participate was made by the country's national counterterrorism headquarters.

Table 1 shows the pre-intervention comparability of participating teams and ports. None of the pre-intervention comparisons suggests a difference between clusters, meaning that randomization appears to have been successful in creating comparable ("balanced") treatment and control groups.

The study was conducted in sea, train, and international ports that handled approximately 1.5 billion passengers during the year prior to the experiment. The 65 teams were comprised of 451 counterterrorism officers who conducted thousands of interviews per year, based on intelligence from counterterrorism agencies and a range of counterterrorism investigation tactics, including behavioral profiling (Dowden et al. 2007; Fox and Farrington 2018). Counterterrorism officers monitor passenger movements and operate searches in cordons in embarkations and disembarkation areas.

Table 1 Comparing participating treatment and control clusters (baseline)

	Control cluster participants	Treatment cluster participants
Number of clusters (teams)	31	34
Clusters per port type:		
Air	22	24
Sea	7	9
Train	2	1
Number of counterterrorism officers	224	227
Mean number of counterterrorism officers per team (SD)	7.22 (3.70)	6.67 (2.47)
Mean number of port terminals per port (SD)	2.03 (1.22)	2.17 (1.36)
Mean number of yearly passengers (in millions) (SD)	24.17 (28.62)	23.87 (29.11)
Mean number of stops per annum per cluster (SD)	1843.14 (1554.27)	1822.33 (1530.52)

These teams, along with security services, are principally responsible for the exercise of counterterrorism powers. To emphasize, these police officers make no other contact with citizens other than dealing directly with suspected terrorist and extremist threats.

Population

Our population consisted of clusters of counterterrorism officers (i.e., teams) who stop and search people suspected of involvement in the commission, preparation, or instigation of acts of terrorism. During the experimental period, 1418 subjects of interest underwent counterterrorism procedures at the participating ports and were included in the study (23.6 subjects per each treatment cluster and 19.8 subjects per control cluster; $\chi^2 = .365$, $p < .10$). Key demographic indicators of the suspects are listed in Table 2. None of the baseline comparisons yielded any clinically significant difference. Running

Table 2 Comparing characteristics of treatment and control cluster suspects interviewed

	Control cluster suspects	Treatment cluster suspects
Number of respondents	615	803
Port type		
Air	51.7%	49.4%
Sea	44.4%	47.3%
Train	3.9%	3.2%
Year of birth (mode)	1983	1982
% male	85.5%	85.8%
% in a relationship	53.5%	52.6%
% with bachelor's degree or higher	31.6%	33.8%
% foreign national	33.7%	37.4%
Self-reported race/ethnicity		
Arab	14.0%	14.0%
Pakistani	14.0%	11.0%
African	4.9%	5.2%
Eastern European	1.3%	2.9%
Indian	1.1%	1.0%
White	21.1%	20.8%
White Irish	14.3%	13.8%
% non-native language	35.3%	37.9%
% follow any religion	63.4%	65.9%
% Christian	28.6%	32.3%
% Muslim	36.8%	35.4%
% reason for travel: business	43.6%	39.0%
% reason for travel: family visit	24.9%	23.8%
% reason for travel: not given	10.9%	12.3%
% first time stopped at the port by police	49.4%	54.6%
% stopped previously by police for any other matter	37.6%	36.2%

statistical significance testing for measurement of baseline differences in randomized controlled trials is not a recommended practice (see de Boer et al. 2015; Harvey 2018; Mills et al. 2019; Moher et al. 2012; Senn 1995).

Random allocation

Randomization was carried out by the research team, not the treatment delivery team. We used simple random assignment using the Cambridge Randomizer (Ariel et al. 2012). Teams within ports were the unit of allocation, with outcomes collected at the level of suspect interviews, clustered within teams. When randomizing teams within ports, one concern was violation of the stable unit treatment value assumption (SUTVA): "non-interference" between units, otherwise described as the outcome of one unit not depending on the outcome of any other unit (see Shadish et al. 2002).

When there is "interference," treatment effects may be attenuated, meaning that the true impact of the independent variable on the dependent variable is masked to some degree (Ariel et al. 2018). All previous experiments on procedural justice, for example, suffered from this problem, which complicates causal inference (Nagin and Sampson 2018). To counter this risk, we randomly assigned entire officer teams in a cluster randomized controlled trial design (Raudenbush 1997). Treatment teams were separated from the control counterterrorism officer teams. Moreover, teams did not engage with one another, did not work on the same cases, had their own assigned line managers, and did not work at the same location as other teams. As such, likelihood of contamination is reduced.

Instrument and the experimental procedure

As reviewed earlier, the literature identifies four dimensions that express "procedural justice" in any interaction between police and the public (Gau 2014; Jonathan-Zamir et al. 2015; Tyler and Wakslak 2004):

1. a citizen's participation in the proceedings prior to an authority reaching a decision ("voice");
2. neutrality of the authority in its decision;
3. whether the authority showed dignity and respect throughout the interaction; and
4. trustworthy motives.

We used one instrument—the PJ Checklist—to operationalize these four dimensions by converting the items into a checklist format (see Supplementary Materials A). Each item corresponded to one of these four elements of procedural justice, and we used multiple items for each dimension in order to improve the likelihood that the construct would in fact be implemented.

The same procedure was mandated in all ports. Before the experiment, the research team communicated with each cluster (i.e., team) and explained to all supervisors the purpose of the experiment and the procedures that were agreed upon with the counterterrorism headquarters (we did not experience any resistance from the local managers). We then met with all the treatment cluster officers, explained the provided guidance on

the PJ Checklist and the rationale behind this project, and consulted with them on the feasibility of the experimental procedures. Modifications were therefore incorporated into the process prior to random assignment. The focus of these meetings was to obtain "buy-in" from the field counterterrorism officers, particularly about the potential benefits of using procedural justice in terms of increasing the perceived legitimacy of the passengers who undergo these interviews.

When counterterrorism officers profiled passengers as potential suspects (based on intelligence briefs or other investigative means), they would ask the suspect to accompany them to an interview room, where the suspect would be questioned with the goal of eliciting information from them. The counterterrorism guidelines in this country dictate that suspects are to be told explicitly that they are being screened for security reasons (as opposed to the possibility of being in possession of contraband, for example). As the process commenced, officers notified their duty supervisor and operations coordinator; the operations coordinators then registered the case on the data portal. The research team audited all case reference numbers against duty logs and found that all cases were delivered as assigned and that no registered case was excluded after treatment delivery.

We distributed paper copies of the PJ Checklists to the counterterrorism officers to use during their interactions with suspects. The implementation of the apparatus was mandated and managed by a local supervising officer in every treatment team; however, supervisors' presence in interviews varied across teams, ranging between 9% and 75% of interviews, with a 31% average presence across all teams. Every interaction was recorded, and a supervisor then signed off on the PJ Checklist completed by the officers. Subsequently, a PJ Checklist policy was applied to all of the treatment subjects, and no PJ Checklists were applied in control conditions (as these teams did not take part in the experimental procedure). We are unable to verify the implementation of the PJ Checklists beyond these self-reported cards and the validations of the line managers. We were not granted access to the interview room and/or any recordings made by the counterterrorism officers; the PJ Checklist formed part of the interview dossier, and we also therefore did not gain access to each PJ Checklist. We discuss this limitation below.

A senior national project led by the country's national counterterrorism headquarters was appointed to serve as the organizational interface with the research team and provide organizational support for the project. In addition, given the distance between the different ports nationwide, we employed local project managers who were vetted to scrutinize and collect data. Their task included creating workable databases that contemporaneously recorded each case that went through the experimental pipeline.

At the end of each interview (both under treatment and control conditions), subjects were handed an envelope by the leading counterterrorism officer, in which was included an invitation to complete a research survey (which had been translated into 11 languages through a professional translation company, including back-translation). As shown in Supplementary Materials B, the survey did not form part of the police interview, and the police officer was not asked to stay with the subjects while they completed the survey. The invitation was always made after the interview was concluded, and per the University of Cambridge's ethical guidelines, consent was requested from all participants prior to the participation in the study; the completion of the survey was voluntary. Once the participant filled out the

survey, they were asked to place the completed survey in a sealed envelope. The duty supervisor collected the sealed envelopes and sent all the surveys to the research team every 2 weeks. The research team then coded the data into Excel for analysis purposes. Over 12 months, 10,225 passengers were suspected of involvement in terrorism and interacted with the counterterrorism officers in the participating ports; of those, 1418 eligible subjects consented to complete the survey—a common level of response rate to surveys in the field of procedural justice research. For example, for mail surveys, Mazerolle et al. (2013a) and Sunshine and Tyler (2003) reported response rates of 15% and 22%, respectively; for telephone surveys, Skogan (2006), Hasisi and Weisburd (2014), Jonathan-Zamir et al. (2016), and Tyler (2006) reported response rates of 35%, 40%, 58%, and 63%, respectively. We found no clinically significant differences between responders and non-responders: The percentage of respondents who are not citizens of this European Democracy in the respondent and non-respondent groups were 41% and 45%, respectively; 19% versus 18% White suspects, respectively; and the mode year of birth was 1982 (SD = 13.1 years) and 1979 (SD = 10.6 years), respectively.

Treatment conditions

During every interrogation, the counterterrorism officer was required to use the PJ Checklist and tick the boxes as they went through the encounter. At the end of the interview, regardless of outcome (detention, arrest, or release), the officer gave the completed PJ Checklist to the duty supervisor, who would then deliver it to the research team.

Control conditions

Control teams applied a "business as usual" approach. This meant that interviews were conducted as they normally would be, without using the PJ Checklist. However, every counterterrorism encounter was recorded by the operations coordinator, and data about the encounter were recorded on the case portal.

Measures

We collected data for several dimensions. The measures were perceptions of distributive justice, effectiveness, obligation to obey the law, willingness to co-operate, and social resistance (for the latter, see Factor et al. 2011, 2013). We collected and report these separately because there is no agreement in the police legitimacy literature on the status of these variables. For example, some scholars operationalize legitimacy as the feeling of obligation the law (Tyler and Fagan 2008), while others view legitimacy as a multidimensional concept, comprising procedural justice, distributive justice, effectiveness and lawfulness (Tankebe 2013 Bottoms and Tankebe 2012). The present experiment does not attempt to "settle the score." Instead, we provide treatment versus control comparisons across multiple measures, as all outcomes are potentially of interest to scholars in this field. Because the relevant theories all hypothesize outcomes in the same direction—improved perceptions of both the specific experience with the counterterrorism officers and global perceptions about counterterrorism and the law

more broadly—the fundamental question then becomes one of magnitude, which we address in the analytical plan below. Overall, six outcome measures were used (but not combined, as they are likely to represent discrete theoretical dimensions).

Distributive justice This variable refers to the fairness of the outcome a person receives in comparison with what others receive (Bottoms and Tankebe 2017). We used four Likert-type items to address this dimension—"The police treated me the same way any other person would have been treated" (Sahin et al. 2017)—on a scale of 1 ("strongly disagree") to 5 ("strongly agree"). The higher the score, the greater the perception of distributive justice. The responses for the items were internally consistent ($\alpha = .91$).

Effectiveness This dimension captures participants' feelings of effectiveness of counterterrorism, measured using multiple items with a Likert-type scale of 1 ("strongly disagree") to 5 ("strongly agree"): "The police are too soft on terrorism" (flipped); "The police are doing a good job in fighting terrorism." The higher the score, the more favorable the judgment of effectiveness. The collapsed dimension was sufficiently reliable ($\alpha = .71$).

Obligation to obey the law This measured participants' sense of duty to obey the law. Three items were used, with responses ranging from 1 ("strongly disagree") to 5 ("strongly agree"). These were combined to create an "obligation-to-obey-law" scale, with a higher score reflecting a greater feeling of obligation ($\alpha = .96$).

Willingness to co-operate with the police This measured participants' intentions to support counterterrorism efforts through the supply of information to the police. Ten questions were used. The response categories were on a scale of 1 to 4. The responses were combined to create a willingness-to-cooperate scale ($\alpha = .98$), with higher scores reflecting a greater intention to cooperate with counterterrorism efforts.

Social resistance This framework offers an addition to the procedural justice model: The notions of alienation and active social resistance that are central to the social resistance framework may play a role in the link between perceptions of procedural justice and non-compliance with the law or engagement in criminal behaviors (Factor et al. 2013). Because of discrimination, members of non-dominant minority groups may feel a lack of attachment to the country and alienation from the larger society. In response, members of these groups may actively engage in counterterrorism, extremism, or other deviant behaviors. By engaging in such behaviors, members of such groups express their willingness and ability to defy the country and the dominant group. To operationalize this framework, we used several items scored on a 1–5 Likert scale: (a) "I often find myself objecting to the symbols of my country (e.g. the flag, the national anthem)"; (b) "I disagree with the values that my country represents"; (c) "sometimes I am opposed to what my country represents"; (d) "it is okay for people who are in a difficult situation to occasionally disobey the law"; (e) "sometimes I get so frustrated I feel like damaging public property"; (f) "sometimes my economic and social status makes me want to show others that I am angry"; and (g) "sometimes I get so frustrated I feel like protesting to express my economic and social status." The items were internally consistent (Cronbach's $\alpha = .87$).

Implementation check

While we are not able to ascertain the extent to which each item in the PJ Checklist was completed, we are, however, in a position to conduct an implementation check by measuring how participants experienced the independent variable—procedural justice—using several questions on a Likert-type scale (see Supplementary Materials B). Each PJ dimension included several items, which were all internally consistent based on Cronbach's alpha (voice ($\alpha = .98$; (Mazerolle et al. 2012)); respect ($\alpha = .97$; (Sahin et al. 2017)); trustworthiness ($\alpha = .98$). Impartiality was measured with a single "flipped" item, which is consistent with other research in this area (Tankebe 2013). For each section of voice, respect, and trustworthiness, higher scores indicate a more favorable perception of the interaction. We used these scales to compare the treatment and control conditions on different dimensions of procedural justice.

In addition, we incorporated specific questions about procedural effectiveness—i.e., the extent to which participants recalled the efficiency of the encounter. Specifically, we asked (a) whether the counterterrorism officer introduced themselves, (b) whether the counterterrorism officer explained what legislation was being used, and (c) the participant's perception of how long it took for the interaction to be completed.

Statistical procedures

Baseline balance First, baseline characteristics, collected at the time of commencing the trial, were cross-tabulated according to the randomized cluster to assess balance and to provide an overview of the study population, both at the cluster and suspect levels (Tables 1 and 2).

Main outcomes We ran a series of linear, logistic, and ordered logit multilevel regression models in Stata version 15.1 to estimate the treatment effect on the dependent variables (Hox et al. 2017).[1] The distribution of the outcome variables dictated the type of model we applied (linear, logistic, or ordered logit). Table 3 reports the raw means and standard deviations by treatment group for each outcome. As all analyses were multilevel models, this accounts for clustering of suspects by location or team, but we also estimated models with robust standard errors. We report exact p values or "< 0.001" where p values were given as "0.000" in raw output. Where multiple outcome testing was undertaken—for example, with procedural justice and subdomains—we report the p values and

---

[1] As Wears (2002) commented, "any time there are multiple analytic options, there will be differences of opinion among statisticians about which choice is best, although most will admit that previous familiarity with a method plays a large role in their opinion. This uncertainty is compounded when there is not a "natural" choice for the analytic unit." (p. 337) We ruled out analyzing the data with the individual participants as the unit of analysis, as many experimenters agree that the allocation unit should be the unit of analysis (Murray 1998).

Table 3 Post-intervention treatment versus control comparison: mean participant responses and dimensions' internal consistency

Dimensions/scales[a]	Control N	Control (SD)[b]	Treatment N	Treatment (SD)	Cronbach's alpha
Procedural justice	597	2.96 (1.53)	766	4.26 (0.97)	.959
Impartiality/neutrality	603	3.05 (1.51)	772	4.19 (1.14)	–[c]
Respect	597	2.97 (1.54)	766	4.26 (0.98)	.968
Trustworthy motives	607	2.92 (1.62)	782	4.25 (1.08)	.984
Voice	603	2.96 (1.58)	772	4.28 (1.00)	.977
Officer introduced himself[d]	572	47%	751	90%	–[c]
Officer explained what legislation is used[c, d]	573	46.3%	760	89.5%	–[c]
Perceived length of interrogation[e]	608	1.95 (1.15)	791	1.94 (1.27)	–[c]
Distributive justice	603	2.98 (1.38)	772	4.13 (0.97)	.914
Effectiveness of counterterrorism policy	561	3.12 (1.23)	726	4.27 (0.91)	.710
Obligation to obey the law	563	2.80 (1.43)	731	4.06 (1.07)	.962
Cooperation with the police	550	2.33 (1.23)	712	3.48 (0.83)	.984
Social resistance	561	2.04 (0.84)	726	1.73 (0.81)	.867

[a] Scale 1–5, except for obligation to obey the law—scale 1–4

[b] SD = standard deviation

[c] single item

[d] Binary

[e] Scale of 1–5; 1–5 scale: (1) < 5 min; (2) 6–15 min; (3) 16–25 min; (4) 26–35 min; (5) 35+ min

include a calculation for the adjusted threshold for statistical significance using the conservative Bonferonni correction (see Abdi 2007; Sedgwick 2012).

To measure the magnitude of the differences between treatment and control conditions, we used two statistics. First, to estimate standardized mean differences and 95% confidence intervals, we used Hedges' *g*, given the unequal cluster sample sizes (Borenstein et al. 2011; Hedges 2007; Nakagawa and Cuthill 2007). The interpretation of Hedges' *g* (Hedges 2007) is the same as Cohen's *d*, meaning these results can be compared with previous police legitimacy experiments (Mazerolle et al. 2013b). We applied this statistic for all outcomes measured in a scale format. Second, we used odds ratios for measures that observed binary outcomes, specifically for the statements "[the] officer introduced himself" and "[the] officer explained what legislation is used."

Results

Main outcomes Table 3 lists the mean participant responses post-intervention across multiple outcomes. There are noticeable differences between treatment and control

clusters, all in the hypothesized direction: The application of a PJ Checklist led to higher scores on the scales for procedural justice, distributive justice, effectiveness, feelings of obligation to obey the law, willingness to cooperate with the police, and lower scores for social resistance. The same pattern emerges for specific items—procedural effectiveness and the various components of procedural justice—when measured independently, where the treatment group expressed more favorable views compared with the control group. For example, while over 90% of the treatment respondents reported that the counterterrorism officer introduced themselves and explained what legislation was used to take the respondent through a counterterrorism interrogation procedure, less than half of control participants reported the same. Of further interest is the perception of time: There were no differences between the treatment and control groups in terms of the perceived length of the interaction with the police officers, which suggests that applying a PJ Checklist does not result in the perception of a longer encounter.

These differences are supported by the statistical results shown in Table 4. Across the board, the independent samples t tests are all highly significant ($p \leq .0001$), and the coefficients are statistically different between groups, with relatively small standard errors. Table 4 further details the magnitude of these differences.

Distributive justice Treatment participants perceived the outcomes they received in comparison with what others received as fairer than control participants ($b = 1.072$ [95% CI 0.818, 1.326]). The effect size was $g = 0.993$ (95% CI 0.880, 1.106).

Effectiveness This dimension captures participants' feelings about the effectiveness of counterterrorism policy. The findings suggest that treatment respondents did not perceive the officers and police as less efficient ($b = 1.046$ [95% CI 0.821, 1.271]). A similar magnitude of difference was detected for this outcome as for the other outcomes ($g = 1.077$ [95% CI 0.959, 1.195]).

In terms of respondents' *willingness to co-operate* with the counterterrorism effort and their *obligation to obey the law*, large differences were reported between groups ($g = 1.022$ [95% CI 0.905, 1.138] and $g = 1.118$ [95% CI 0.999, 1.238], respectively). Read together with the group-level mean differences (Table 3), while the mean score for the control group ranged from 56 to 58, the treatment group's mean scores range from 81 to 87 (on a scale of 1–100)—more than one standard deviation apart.

Social resistance PJ Checklists reduced the expressed social resistance and willingness of treatment participants to engage in criminal or deviant behavior compared with control conditions ($b = -0.308$ [95% CI $-0.519, -0.096$]), but had the weakest effect size ($g = -0.370$ [95% CI $-0.259, -0.482$]). This result is probably due to the fact that respondents were less likely to express perceptions such as "I often find myself objecting to the symbols of my country" or to disagree with the core values that their home countries represent (see Table 3). Still, the prevalence of these resistance perceptions was not negligible, with more than a third of respondents expressing these views.

Table 4 Effect of PJ Checklist on suspected terrorists' ratings of legitimacy-related dimensions: regression coefficients and effect sizes from multi-level linear, logistic, and ordered logit regression models (N=65 clusters)

Variable	Coefficient	95% CI lower	95% CI upper	SE[a]	p value	ICC	Sample size	Effect size (95% CI)
Procedural justice	1.32	1.04	1.60	0.14	< 0.001[b]	.108	1363	g = 1.044 (0.930, 1.158)
Impartiality/neutrality	1.16	0.91	1.42	0.13	< 0.001[b]	.068	1375	g = 1.031 (0.918, 1.145)
Respect	1.31	1.02	1.60	0.15	< 0.001[b]	.114	1363	g = 1.024 (0.910, 1.137)
Trustworthy motives	1.36	1.07	1.65	0.15	< 0.001[b]	.091	1389	g = 1.001 (0.888, 1.113)
Voice	1.34	1.06	1.63	0.15	< 0.001[b]	.104	1375	g = 0.873 (0.761, 0.984)
Officer introduced himself [c]	2.71	2.12	3.29	0.30	< 0.001	.175	1323	OR[d] = 14.97 (8.35, 26.86)
Officer explained what legislation is used[c]	2.54	1.96	3.11	0.29	< 0.001	.170	1333	OR[d] = 12.62 (7.08, 22.50)
Perceived length of interrogation[e]	− 0.22	− 0.66	0.22	0.22	0.326	n/a	1400	OR[d] = 0.80 (0.52, 1.24)
Distributive justice	1.16	0.92	1.40	0.12	< 0.001	.095	1375	g = 0.993 (0.880, 1.106)
Effectiveness	1.20	0.97	1.43	0.12	< 0.001	.100	1287	g = 1.077 (0.959, 1.195)
Obligation to obey the law	1.35	1.08	1.62	0.14	< 0.001	.108	1294	g = .022 (0.905, 1.138)
Cooperation with the police	1.21	0.98	1.44	0.12	< 0.001	.118	1262	g = 1.118 (0.999, 1.238)
Social resistance	− 0.26	− 0.42	− 0.11	0.08	< 0.001	.068	1287	g = − 0.370 (− 0.259, − 0.482)

[a] Adjusted for clustering and using robust SEs

[b] Bonferonni threshold for "significance" adjusted for five tests would be 0.01—all reported p values < 0.001

[c] Logistic regression coefficient

[d] OR = odds ratio

[e] Ordered logit model

Perceptions of procedural justice PJ Checklists led to improved perceptions of procedural justice (b = 1.272 [95% CI 1.019, 1.525]), with a large effect size (g = 1.044 [95% CI 0.930, 1.158]). The same findings and effect sizes were found for each of the four dimensions that comprise this scale—voice, trustworthy motives, respect, and impartiality—as well as for the procedural effectiveness variables. When using PJ Checklists, officers are significantly more likely to introduce themselves to suspects and to inform suspects of their legal status.

Discussion

The secrecy that characterizes counterterrorism limits the scope of experimental research in this area of law enforcement (Lum et al. 2006). As noted previously, public safety concerns in which valid estimates of intervention effects could prove beneficial are precisely the situations in which evidence is scarce; this evidence paradox obstructs any attempt to robustly test the effectiveness of these policies. However, this study was able to produce some of the first rigorous evidence regarding whether it is possible to

change the behavior of officers conducting counterterrorism interviews. This is a crucial turning point in the development of evidence-led policymaking in counter-terrorism (Sanderson 2002; Sherman 2013). Our experiment illustrates that rigorous impact evaluations in this extremely sensitive area of criminal justice policy are possible, and we hope that replications and follow-up research will be conducted in the future to further assess the results presented here.

Procedural justice is a key component of police legitimacy (Bottoms and Tankebe 2012, 2017). We have shown here that PJ checklists offer a simple, scalable means of improving how state agents interact with suspected terrorists. The police can enhance people's obligation to obey counterterrorism laws and improve the flow of information, even in a counterterrorism environment, with what is seemingly a cost-effective tool: a checklist that takes very little time to complete. A straightforward reminder to treat suspects in accordance with the tenets of fairness, dignity, equality, and respect has resulted in large effects, in the magnitude of more than an entire standard deviation from the control group. We found improvements in every dimension measured. Specifically, terrorism suspects were given more information about their interrogation, were allowed to voice their concerns, were treated equally in comparison to other suspects, and reported higher overall legitimacy scores compared with control conditions. The checklist altered how suspects perceive both the "local" interaction with specific counterterrorism officers and "global" dimensions of counterterrorism policies.

Previous studies that sought to increase police legitimacy resulted in relatively modest effect sizes for procedural justice (Mazerolle et al. 2013b). The effect sizes reported in the present paper are substantially larger. This difference can be interpreted in two ways: methodological and theoretical. To our knowledge, this is the first cluster randomized trial in this area of research. Using a cluster-randomized design overcomes many of the threats to internal validity that occur in other randomized controlled trial designs (Ariel et al. 2018). For example, minimizing contamination through design means that we have avoided problems that have affected other PJ experiments—e.g., the same officers being in both treatment and control conditions (Langley 2014; Mazerolle, Antrobus, et al., 2013). Through minimizing this particular threat to validity, a cleaner treatment/control comparison is possible.

Second, and more conceptually (but with a clear implication for public policy), procedural justice is important in every type of law-enforcement procedure (Nagin and Telep 2017a; Tyler et al. 2014), specifically when dealing with individuals suspected of terrorism. Participants in this study experienced what must have been an emotionally and psychologically intense event. In the control group, the overall scores were far from "positive perceptions." For example, as shown in Table 3, the non–PJ Checklist group expressed overall negative views; without the checklist, most suspects did not perceive the interrogation as fair or legitimate. We can only speculate as to how such negative experiences with the police are reported in the interviewee's community, thereby perpetuating a vicious cycle of perceived illegitimacy (Skogan 2006). However, once the officers were "nudged" to be more procedurally just, suspects subsequently reported these interactions as more fair, just, and effective.

Democracies have the inalienable right to defend themselves to the fullest extent of the law. At the same time, however, even in the context of counterterrorism and extremism, societies should not look the other way when fundamental human and legal rights risk being sidetracked in the name of safety and security (Zedner 2003). As the

police often rely on the public to report threats or cooperate with counterterrorism efforts, the police need to be viewed as wielding their state-bestowed powers legitimately (Hough et al. 2010). As such, it seems sensible to conduct interviews with terrorism suspects in a way that engenders the highest possible levels of perceived fairness and procedurally just conduct. Failing to do so means that the rights of innocent citizens may be routinely violated (Luban 2002), which may in turn affect wider community perceptions of police and policing legitimacy (Tyler 2012). While we have no evidence that a PJ Checklist policy leads to more general legitimacy, manifested as more cooperation or less extremism (see Nagin and Telep 2017b), this does not mean that the goal of changing perception is unimportant. [2] Legitimacy is one of the "ultimate values" by which policing is judged, and it creates a potential "cushion of support" for police in times of difficulty, such as terrorist attacks (Rasinski et al. 1985). If treated unfairly, why would communities cooperate? If counterterrorism interrogations apply extreme measures and violate human rights, who would dare report a suspect? These very tactics can drive some elements into extremism, in what can only be viewed as a backfire of a counterterrorism policy (Chalk 2017). Our study shows that a simple checklist may make a disproportionately large contribution to enhancing a range of perception associated with legitimacy in one of the earliest phases of the counterterrorism interview process.

Given the evidence presented here, a practical policy implication can be identified: incorporating a PJ Checklist should be part and parcel of counterterrorism policing's standard operating procedure. The approach is inexpensive, and the PJ Checklist can be adapted for and tested in other jurisdictions, practices, or needs. We would also argue that if checklists are useful for counterterrorism, they are likely to be valuable in other contentious law-enforcement encounters: stop-and-accounts, searches, and criminal interviews more broadly. However, replications, preferably in diverse samples, are still required to confirm this.

At the same time, if officers will not adhere to the code of practice which serves as the foundation for the PJ Checklist, and if line managers will not interject when the checklists do not serve their purpose, the PJ Checklist policy will be "toothless" (Ariel 2012). An organizationally-led implementation policy is necessary for the rules to be followed systematically and continuously, with clear guidelines for taking steps to deal with non-compliance (Fixsen et al. 2009).

A PJ Checklist is not a salve for egregious violations of suspects' human rights. In places where outright torture is practiced, a policy of ticking boxes will not matter much. However, our study takes place in a European democracy that, broadly speaking, has a demonstrated record of securing human rights, like most of the developed The Organization for Economic Co-operation and Development's (OECD) countries. It remains true that, in similar countries, a constant clash exists between security needs and securing human rights (Hasisi and Weisburd 2011; Merin et al. 2015; Newheiser and DeMarco 2018). The counterterrorism code of practice expresses this dilemma, but it still requires that suspects be treated fairly. It turns out that officers need to be reminded to follow these

---

[2] We concluded that the treatment measures the effectiveness of counterterrorism measures; however, we had no measures on the security outcomes—i.e., the false positive or negative identification of terrorists—resulting from the interrogations. We did not have access to these data points from the security services. Future research should attempt to provide non-perceptual consequences of the treatment effects.

rules, just like pilots, surgeons, and engineers. Therefore, for security agencies that express a genuine concern for due process, a little checklist can go a long way.

Notwithstanding these findings, we are cognizant that one of, if not the most, important pieces of the puzzle is missing: We did not have the opportunity to test actual cooperation with the police in its counterterrorism mission (see Nagin and Telep 2017a, as well as The National Academies of Sciences, Engineering, and Medicine 2018, on proactive policing for a broader critique). We have at least been able to show improvements in the *willingness* of those who interact with counterterrorism officers to act upon the tested reflections. Differences from control participants were not negligible, and treatment participants expressed a greater willingness to report various terrorism-related activities to the police, including different threats to national security. As important as these perceptions are in democratic societies (Tyler 2017), and notwithstanding the view that procedural justice ought to be practiced inconsequentially to the question of tangible benefits (Nagin and Telep 2017b), there is no evidence linking the PJ Checklist to *behavioral* manifestations. We did not gain access to operational intelligence data that links those who underwent a PJ Checklist treatment (or control conditions) to future reporting of any crime to the police or the counterterrorism information hotline. Future controlled tests in police legitimacy research ought to delve into these potential benefits in greater depth.

References

Abdi, H. (2007). The Bonferroni and Šidák corrections for multiple comparisons. In N. J. Salkind & K. Rasmussen (Eds.), *Encyclopedia of measurement and statistics* (Vol. 3, pp. 103–107). SAGE Publications.

Ajzenstadt, M., & Ariel, B. (2008). Terrorism and risk management: The Israeli case. *Punishment & Society: International Journal of Penology, 10*(4), 355–374.

Amnesty International & Open Society Foundations. (2017). *Inhuman and unnecessary: Human rights violations in Dutch high-security prisons in the context of counterterrorism*.

Andrighetto, G., Paolucci, M., Paus, A., Szekely, A., Trussardi, P., Calderoni, F., Campedelli, G. M., Comunale, T., Frualdo, N., Furfaro, E., Weisburd, D., Hasisi, B., & Wolfowicz, M. (2019). *Development of agent based simulations of OCTN*. PROTON. https://www.researchgate.net/profile/FrancescoCalderoni/publication/336770200 D51 Development of Agent Based Simulations of organised crime and terrorist networks/links/5db15c2ba6fdccc99d939667/D51-Development-of-Agent-Based-Simulations-of-organised-crime-and-terrorist-networks.pdf.

Antrobus, E., Thompson, I., & Ariel, B. (2018). Procedural justice training for police recruits: results of a randomized controlled trial. *Journal of Experimental Criminology, 15*(1), 29–53.

Ariel, B. (2012). Deterrence and moral persuasion effects on corporate tax compliance: findings from a randomized controlled trial. *Criminology, 50*(1), 27–69.

Ariel, B., Vila, J., & Sherman, L. (2012). Random assignment without tears: how to stop worrying and love the Cambridge randomizer. *Journal of Experimental Criminology, 8*(2), 193–208.

Ariel, B., Sutherland, A., & Sherman, L. W. (2018). Preventing treatment spillover contamination in criminological field experiments: the case of body-worn police cameras. *Journal of Experimental Criminology, 15*(4), 569–591.

Ariel, B., Mitchell, R. J., Tankebe, J., Firpo, M. E., Fraiman, R., & Hyatt, J. M. (2020). Using wearable technology to increase police legitimacy in Uruguay: the case of body-worn cameras. *Law & Social Inquiry, 45*(1), 52–80.

Bamford, B. W. C. (2005). The role and effectiveness of intelligence in Northern Ireland. *Intelligence and National Security, 20*(4), 581–607.

Bayley, D. H., & Weisburd, D. (2009). Cops and spooks: the role of police in counterterrorism. In D. Weisburd, T. Feucht, I. Hakimi, L. Mock, & S. Perry (Eds.), *To protect and to serve: policing in an age of terrorism* (pp. 81–99). Berlin: Springer.

Bonner, D. (2013). Counter-terrorism and European human rights since 9/11: The United Kingdom experience. *European Public Law, 19*(1), 97–128.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Hoboken: Wiley.

Bottoms, A., & Tankebe, J. (2012). Beyond procedural justice: a dialogic approach to legitimacy in criminal justice. *The Journal of Criminal Law and Criminology, 102*(1), 119–170.

Bottoms, A., & Tankebe, J. (2017). *Police legitimacy and the authority of the state*. Oxford: Hart Publishing Limited.

Brecher, B., & Brecher, R. (2007). *Torture and the ticking bomb*. Hoboken: Wiley.

Byman, D. (2015). The homecomings: what happens when Arab foreign fighters in Iraq and Syria return? *Studies in Conflict & Terrorism, 38*(8), 581–602.

Chalk, P. (2017). The response to terrorism as a threat to liberal democracy. In *War on terrorism* (pp. 83–98). Abingdon: Routledge.

Chang, N. (2011). *Silencing political dissent: how post-September 11 anti-terrorism measures threaten our civil liberties*. New York: Seven Stories Press.

Clayton, R., Tomlinson, H., & George, C. (2000). *The law of human rights* (Vol. 2). Oxford: Oxford University Press.

Cronin, A. K. (2006). How Al-Qaida ends: the decline and demise of terrorist groups. *International Security, 31*(1), 7–48.

Curzer, H. J. (2006). Admirable immorality, dirty hands, ticking bombs, and torturing innocents. *The Southern Journal of Philosophy, 44*(1), 31–56.

Darmer, M. K. B. (2002). Beyond Bin Laden and Lindh: confessions law in an age of terrorism. *Cornell Journal of Law and Public Policy, 12*(2), 319–372.

Davis, D. W., & Silver, B. D. (2004). Civil liberties vs. security: public opinion in the context of the terrorist attacks on America. *American Journal of Political Science, 48*(1), 28–46.

de Boer, M. R., Waterlander, W. E., Kuijper, L. D., Steenhuis, I. H., & Twisk, J. W. (2015). Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity, 12*(1), 4.

Degani, A., & Wiener, E. L. (1993). Cockpit checklists: concepts, design, and use. *Human Factors, 35*(2), 345–359.

Demir, M., Apel, R., Braga, A. A., Brunson, R. K., & Ariel, B. (2018). Body worn cameras, procedural justice, and police legitimacy: a controlled experimental evaluation of traffic stops. *Justice Quarterly, 37*(1), 1–32.

Dershowitz, A. M. (2002). Want to torture? Get a warrant. San Francisco Chronicle.

Dotan, Y. (1999). Judicial rhetoric, government lawyers, and human rights: the case of the Israeli high court of justice during the intifada. *Law and Society Review, 33*(2), 319–363.

Dotan, Y. (2004). Legalising the unlegaliseable: terrorism, secret services and judicial review in Israel 1970–2001. *Judicial Review and Bureaucratic Impact*, 190–220.

Dowden, C., Bennell, C., & Bloomfield, S. (2007). Advances in offender profiling: a systematic review of the profiling literature published over the past three decades. *Journal of Police and Criminal Psychology, 22*(1), 44–56.

Dreher, A., Gassebner, M., & Siemers, L. H. (2010). Does terrorism threaten human rights? Evidence from panel data. *The Journal of Law and Economics. 53*(1), 65–93.

Drizin, S. A., & Colgan, B. A. (2000). Let the cameras roll: mandatory videotaping of interrogations is the solution to Illinois' problem of false confessions. *Loyola University Chicago Law Journal, 32*, 337–424.

Europol. (2017). European Union terrorism situation and trend report 2017. Available at https://www.europol.europa.eu/activities-services/main-reports/eu-terrorism-situation-and-trend-report-te-sat-2017.

Factor, R., Kawachi, I., & Williams, D. R. (2011). Understanding high-risk behavior among non-dominant minorities: a social resistance framework. *Social Science & Medicine, 73*(9), 1292–1301.

Factor, R., Mahalel, D., Rafaeli, A., & Williams, D. R. (2013). A social resistance perspective for delinquent behavior among non-dominant minority groups. *British Journal of Criminology, 53*(5), 784–804.

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice, 19*(5), 531–540.

Fox, B., & Farrington, D. P. (2018). What have we learned from offender profiling? A systematic review and meta-analysis of 40 years of research. *Psychological Bulletin, 144*(12), 1247.

Gau, J. M. (2014). Procedural justice and police legitimacy: a test of measurement and structure. *American Journal of Criminal Justice, 39*(2), 187–205.

Gawande, A. (2009). The checklist manifesto: How to get things right. Profile.

Ginbar, Y. (2008). *Why not torture terrorists?: moral, practical and legal aspects of the "ticking bomb" justification for torture*. Oxford: Oxford University Press.

Godsey, M. A. (2005). Rethinking the involuntary confession rule: toward a workable test for identifying compelled self-incrimination. *California Law Review, 93*, 465–540.

Gordis, L. (2000). More on causal inferences: bias, confounding, and interaction. *Epidemiology, 4*, 247–264.

Grant, E. (2007). Dignity and equality. *Human Rights Law Review, 7*, 299.

Gronke, P., Rejali, D., Drenguis, D., Hicks, J., Miller, P., & Nakayama, B. (2010). U.S. public opinion on torture, 2001–2009. *Political Science & Politics, 43*(3), 437–444.

Gross, E. (2001). Democracy in the war against terrorism—the Israeli experience. *Loyola of Los Angeles Law Review, 35*, 1161.

Gross, E. (2004). The struggle of a democracy against terrorism - protection of human rights: the right to privacy versus the national interest - the proper balance. *Cornell International Law Journal, 37*, 27.

Guiora, A. N. (2005). Legislative and policy responses to terrorism, a global perspective. *San Diego International Law Journal, 7*, 125.

Guiora, A. N. (2006). Where are terrorists to be tried: a comparative analysis of rights granted to suspected terrorists. *Catholic University Law Review, 56*, 805.

Hales, B. M., & Pronovost, P. J. (2006). The checklist—a tool for error management and performance improvement. *Journal of Critical Care, 21*(3), 231–235.

Harvey, L. A. (2018). Statistical testing for baseline differences between randomized groups is not meaningful. *Spinal Cord, 56*, 919.

Hasisi, B., & Weisburd, D. (2011). Going beyond ascribed identities: the importance of procedural justice in airport security screening in Israel. *Law and Society Review, 45*(4), 867–892.

Hasisi, B., Alpert, G. P., & Flynn, D. (2009). The impacts of policing terrorism on society: lessons from Israel and the U.S. In D. Weisburd, T. Feucht, I. Hakimi, L. Mock, & S. Perry (Eds.), *To Protect and to serve: Policing in an age of terrorism* (pp. 177–202). Berlin: Springer.

Hasisi, B., Perry, S., Ilan, Y., & Wolfowicz, M. (2019). Concentrated and close to home: the spatial clustering and distance decay of lone terrorist vehicular attacks. *Journal of Quantitative Criminology*, 1–39. https://doi.org/10.1007/s10940-019-09414-z.

Heath-Kelly, C. (2012). Reinventing prevention or exposing the gap? False positives in U.K. terrorism governance and the quest for pre-emption. *Critical Studies on Terrorism, 5*(1), 69–87.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*(4), 341–370.

Hewitt, C. (1990). Terrorism and public opinion: a five country comparison. *Terrorism and Political Violence, 2*(2), 145–170.

Hirose, I. (2014). *Egalitarianism*. Abingdon: Routledge.

Hood, R., & Hoyle, C. (2015). *The death penalty: a worldwide perspective*. Oxford: Oxford University Press.

Hough, M., Jackson, J., Bradford, B., Myhill, A., & Quinton, P. (2010). Procedural justice, trust, and institutional legitimacy. *Policing: A Journal of Policy and Practice, 4*(3), 203–210.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: techniques and applications*. Abingdon: Routledge.

Huis, A., van Achterberg, T., de Bruin, M., Grol, R., Schoonhoven, L., & Hulscher, M. (2012). A systematic review of hand hygiene improvement strategies: a behavioral approach. *Implementation Science, 7*(1), 92.

Huq, A. Z., Tyler, T. R., & Schulhofer, S. J. (2011). Mechanisms for eliciting cooperation in counterterrorism policing: evidence from the United Kingdom. *Journal of Empirical Legal Studies, 8*(4), 728–761.

Hyatt, J. M., Mitchell, R. J., & Ariel, B. (2017). The effects of a mandatory body-worn camera, policy on officer perceptions of accountability, oversight, and departmental culture. *Villanova Law Review, 62*, 1005.

Jonas, J., & Harper, J. (2006). *Effective counterterrorism and the limited role of predictive data mining*. Washington, D.C.: Cato Institute.

Jonathan-Zamir, T., & Harpaz, A. (2018). Predicting support for procedurally just treatment: the case of the Israel national police. *Criminal Justice and Behavior, 45*(6), 840–862.

Jonathan-Zamir, T., Mastrofski, S. D., & Moyal, S. (2015). Measuring procedural justice in police-citizen encounters. *Justice Quarterly, 32*(5), 845–871.

Jonathan-Zamir, T., Hasisi, B., & Margalioth, Y. (2016). Is it the what or the how? The roles of high-policing tactics and procedural justice in predicting perceptions of hostile treatment: the case of security checks at ben-Gurion airport, Israel. *Law and Society Review, 50*(3), 608–636.

Kreps, S. (2014). Flying under the radar: a study of public attitudes towards unmanned aerial vehicles. *Research & Politics, 1*(1), 1–14.

Kugler, M. B., & Cooper, J. (2010). Still an American? Mortality salience and treatment of suspected terrorists. *Journal of Applied Social Psychology, 40*(12), 3130–3147.

LaFree, G. (2018). *Losing legitimacy: street crime and the decline of social institutions in America*. Abingdon: Routledge.

Langley, B. (2014). *A randomized control trial comparing the effects of procedural justice to experienced utility theories in airport security stops* [unpublished master's dissertation]. Cambridge: University of Cambridge.

Laycock, G. (2012). In support of evidence-based approaches: a response to Lum and Kennedy. *Policing: A Journal of Policy and Practice, 6*(4), 324–326.

Leslie, I. (2017). *The scientists persuading terrorists to spill their secrets*. Kings Place: The Guardian.

Lindekilde, L., & Sedgwick, M. (2012). *Impact of counter-terrorism on communities: Denmark background report*. Institute for Strategic Dialogue. Available at https://s3.amazonaws.com/academia.edu.documents/30871721/Country_report_Denmark_AD_15_Oct_forwebsite.pdf?response-content-disposition=inline%3B%20filename%3DImpact_of_Counter-Terrorism_on_Communiti.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20200308%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20200308T175540Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=6a661c2b5a345b3e04655aea2b92db8dae82d6feecbede743951051245b2241f.

Luban, D. (2002). The war on terrorism and the end of human rights. *Philosophy & Public Policy Quarterly, 22*(3), 9–14.

Lum, C., & Kennedy, L. W. (2012). In support of evidence-based approaches: a rebuttal to Gloria Laycock. *Policing: A Journal of Policy and Practice, 6*(4), 317–323.

Lum, C., Kennedy, L. W., & Sherley, A. (2006). Are counter-terrorism strategies effective? The results of the Campbell systematic review on counter-terrorism evaluation research. *Journal of Experimental Criminology, 2*(4), 489–516.

Lum, C., Crafton, P. Z., Parsons, R., Beech, D., Smarr, T., & Connors, M. (2015). Discretion and fairness in airport security screening. *Security Journal, 28*(4), 352–373.

MacQueen, S., & Bradford, B. (2015). Enhancing public trust and police legitimacy during road traffic encounters: results from a randomized controlled trial in Scotland. *Journal of Experimental Criminology, 11*(3), 419–443.

MacQueen, S., & Bradford, B. (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), 321-345.

Malloy, L. C., Shulman, E. P., & Cauffman, E. (2014). Interrogations, confessions, and guilty pleas among serious adolescent offenders. *Law and Human Behavior, 38*(2), 181.

Marshburn, T., Whitmore, M., Ortiz, R., Segal, M., Smart, K., & Hughes, C. (2003). An independent human factors analysis and evaluation of the emergency medical protocol checklist for the International Space Station. Available at https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110003477.pdf

Mastrofski, S. D., Snipes, J. B., & Supina, A. E. (1996). Compliance on demand: the public's response to specific police requests. *Journal of Research in Crime and Delinquency, 33*(3), 269–305.

Mazerolle, L., Bennett, S., Antrobus, E., & Eggins, E. (2012). Procedural justice, routine encounters and citizen perceptions of police: main findings from the Queensland community engagement trial (Qcet). *Journal of Experimental Criminology, 8*(4), 343–367.

Mazerolle, L., Antrobus, E., Bennett, S., & Tyler, T. R. (2013a). Shaping citizen perceptions of police legitimacy: a randomized field trial of procedural justice. *Criminology, 51*(1), 33–63.

Mazerolle, L., Bennett, S., Davis, J., Sargeant, E., & Manning, M. (2013b). Procedural justice and police legitimacy: a systematic review of the research evidence. *Journal of Experimental Criminology, 9*(3), 245–274.

McCluskey, J. D. (2003). *Police requests for compliance: coercive and procedurally just tactics*. New York: LFB Scholarly Publishing LLC.

McCrudden, C. (2008). Human dignity and judicial interpretation of human rights. *European Journal of International Law, 19*(4), 655–724.

Merin, O., Goldberg, S., & Steinberg, A. (2015). Treating terrorists and victims: a moral dilemma. *The Lancet, 385*(9975), 1289.

Mills, L. G., Barocas, B., Butters, R. P., & Ariel, B. (2019). A randomized controlled trial of restorative justice-informed treatment for domestic violence crimes. *Nature Human Behavior, 3*(12), 1284–1294.

Mitchell, R. J., Ariel, B., Firpo, M. E., Fraiman, R., del Castillo, F., Hyatt, J. M., Weinborn, C., & Sabo, H. B. (2018). Measuring the effect of body-worn cameras on complaints in Latin America: the case of traffic police in Uruguay. *Policing: An International Journal, 41*(4), 518–524.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., … & Altman, D. G. (2012). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomized trials. *International Journal of Surgery, 10*(1), 28–55.

Moran, J. (2005). State power in the war on terror: a comparative analysis of the UK and USA. *Crime, Law and Social Change, 44*(4–5), 335–359.

Mueller, J. E., & Stewart, M. G. (2016). *Chasing ghosts: the policing of terrorism*. Oxford: Oxford University Press.

Murphy, K., Hinds, L., & Fleming, J. (2008). Encouraging public cooperation and support for police. *Policing and Society, 18*(2), 136–155.

Murphy, K., Mazerolle, L., & Bennett, S. (2014). Promoting trust in police: findings from a randomized experimental field trial of procedural justice policing. *Policing and Society, 24*(4), 405–424.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. Oxford: Oxford University Press.

Nagin, D. S., & Sampson, R. J. (2018). The real gold standard: Measuring counterfactual worlds that matter most to social science and policy. In *Annual Review of Criminology* (Vol. 2, pp. 123–145).

Nagin, D. S., & Telep, C. W. (2017a). Procedural justice and legal compliance. *Annual Review of Law and Social Science, 13*, 5–28.

Nagin, D. S., & Telep, C. W. (2017b). Response to "procedural justice and policing: a rush to judgment?". *Annual Review of Law and Social Science, 13*, 55–58.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews, 82*(4), 591–605.

National Academies of Sciences, Engineering, and Medicine. (2018). *Proactive policing: effects on crime and communities*. Washington, D.C.: National Academies Press.

Newheiser, A.-K., & DeMarco, T. C. (2018). Who deserves basic rights? People condone violations of procedural and physical rights in the treatment of terrorist suspects. *Law and Human Behavior, 42*(1), 50–56.

Obama, B. (2010). *National security strategy of the United States*. Darby: Diane Publishing.

Office for Security and Counter-Terrorism. (2012). *Review of the operation of Schedule 7: A public consultation*. Home Office. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/212548/WEB_-_2013_07_15_Review_of_the_operation_of_Schedule_7_A_Public_Consulta___.pdf.

Office of the United Nations High Commissioner for Human Rights. (2008). *Human rights, terrorism and counter-terrorism*. United Nations. Available at https://www.ohchr.org/Documents/Publications/Factsheet32EN.pdf.

Perito, R., & Parvez, T. (2014). *A counterterrorism role for Pakistan's police stations*. Washington, D.C.: U.S. Institute of Peace.

Perliger, A., Hasisi, B., & Pedahzur, A. (2009). Policing terrorism in Israel. *Criminal Justice and Behavior, 36*(12), 1279–1304.

Perry, G., & Hasisi, B. (2018). Closing the gap: promoting suspect communities' cooperation with airport security. *Terrorism and Political Violence*, 1–20. https://doi.org/10.1080/09546553.2018.1442331.

Perry, S., & Jonathan-Zamir, T. (2014). Lessons from empirical research on policing in Israel: policing terrorism and police–community relationships. *Police Practice and Research, 15*(2), 173–187.

Piazza, J. A., & Walsh, J. I. (2010). Physical integrity rights and terrorism. *Political Science & Politics, 43*(3), 411–414.

Rae, J. A. (2012). Will it ever be possible to profile the terrorist? *Journal of Terrorism Research, 3*(2). https://doi.org/10.15664/jtr.380.

Randahl, D. (2018). Terrorism and public opinion: the effects of terrorist attacks on the popularity of the president of the United States. *Terrorism and Political Violence, 30*(3), 373–383.

Rasinski, K., Tyler, T. R., & Fridkin, K. (1985). Exploring the function of legitimacy: mediating effects of personal and institutional legitimacy on leadership endorsement and system support. *Journal of Personality and Social Psychology, 49*(2), 386–394.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*(2), 173.

Rehman, J. (2007). Islam, war on terror and the future of Muslim minorities in the United Kingdom: dilemmas of multiculturalism in the aftermath of the London bombings. *Human Rights Quarterly, 29*, 831–878.

Reiss, A. J. (1971). *The police and the public*. New Haven: Yale University Press.

Reiss Jr., A. J., & Farrington, D. P. (1991). Advancing knowledge about co-offending: results from a prospective longitudinal survey of London males. *The Journal of Criminal Law and Criminology, 82*, 360.

Rothe, D., & Muzzatti, S. L. (2004). Enemies everywhere: terrorism, moral panic, and U.S. civil society. *Critical Criminology, 12*(3), 327–350.

Sahin, N., Braga, A. A., Apel, R., & Brunson, R. K. (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), 701–726.

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration, 80*(1), 1–22.

Sangero, B. (2016). *Safety from false convictions*. Available at SSRN 2816292.

Schmid, A. P. (2013). Radicalisation, de-radicalisation, and counter-radicalisation: a conceptual discussion and literature review. *ICCT Research Paper, 97*(1), 22.

Sedgwick, P. (2012). Multiple significance tests: the Bonferroni correction. *BMJ, 344*, e509.

Senn, S. J. (1995). Base logic: tests of baseline balance in randomized clinical trials. *Clinical Research and Regulatory Affairs, 12*(3), 171–182.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shane, S. (2009). Waterboarding used 266 times on 2 suspects. *New York Times*.

Sherman, L. W. (2013). The rise of evidence-based policing: targeting, testing, and tracking. *Crime and Justice, 42*(1), 377–451.

Sindhav, B., Holland, J., Rodie, A. R., Adidam, P. T., & Pol, L. G. (2006). The impact of perceived fairness on satisfaction: are airport security measures fair? Does it matter? *Journal of Marketing Theory and Practice, 14*(4), 323–335.

Skogan, W. G. (2006). Asymmetry in the impact of encounters with police. *Policing and Society, 16*(02), 99–126.

Spinney, L. (2010). Science in court: the fine print. *Nature News, 464*(7287), 344–346.

Stewart, M. G., & Mueller, J. (2013). Terrorism risks and cost-benefit analysis of aviation security. *Risk Analysis, 33*(5), 893–908.

Steyn, J. (2004). Guantanamo Bay: the legal black hole. *International & Comparative Law Quarterly, 53*(1), 1–15.

Sunshine, J., & Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law and Society Review, 37*(3), 513–548.

Tankebe, J. (2013). Policing, procedural fairness and public behavior: a review and critique. *International Journal of Police Science and Management, 11*(1), 8–19.

Treadwell, J. R., Lucas, S., & Tsou, A. Y. (2014). Surgical checklists: a systematic review of impacts and implementation. *BMJ Quality and Safety, 23*(4), 299–318.

Tyler, T. R. (2006). *Why people obey the law*. Princeton: Princeton University Press.

Tyler, T. R. (2012). Toughness vs. fairness: police policies and practices for managing the risk of terrorism. In C. Lum, L. W. Kennedy, & L. W (Eds.), *Evidence-based counterterrorism policy* (pp. 353–363). Berlin: Springer.

Tyler, T. R. (2017). Procedural justice and policing: a rush to judgment? *Annual Review of Law and Social Science, 13*, 29–53.

Tyler, T. R., & Bies, R. J. (1990). Beyond formal procedures: the interpersonal context of procedural justice. *Applied Social Psychology and Organizational Settings, 77*, 98.

Tyler, T. R., & Fagan, J. (2008). Legitimacy and cooperation: why do people help the police fight crime in their communities. *Ohio State Journal of Criminal Law, 6*, 231.

Tyler, T. R., & Wakslak, C. J. (2004). Profiling and police legitimacy: procedural justice, attributions of motive, and acceptance of police authority. *Criminology, 42*(2), 253–282.

Tyler, T. R., Schulhofer, S., & Huq, A. Z. (2010). Legitimacy and deterrence effects in counterterrorism policing: a study of Muslim Americans. *Law and Society Review, 44*(2), 365–402.

Tyler, T. R., Fagan, J., & Geller, A. (2014). Street stops and police legitimacy: teachable moments in young urban men's legal socialization. *Journal of Empirical Legal Studies, 11*(4), 751–785.

Viscusi, W. K., & Zeckhauser, R. J. (2003). Sacrificing civil liberties to reduce terrorism risks. *Journal of Risk and Uncertainty, 26*(2–3), 99–120.

Wainer, H., & Savage, S. (2008). Until proven guilty: false positives and the war on terror. *Chance, 21*(1), 55–58.

Wallace, D., White, M. D., Gaub, J. E., & Todak, N. (2018). Body-worn cameras as a potential source of depolicing: testing for camera-induced passivity. *Criminology, 56*(3), 481–509.

Watson, G. (2019). *Respect and criminal justice*. Oxford: Clarendon Press.

Wears, R. L. (2002). Advanced statistics: statistical methods for analyzing cluster and cluster-randomized data. *Academic Emergency Medicine, 9*(4), 330–341.

Weiser, T. G., Haynes, A. B., Lashoher, A., Dziekan, G., Boorman, D. J., Berry, W. R., & Gawande, A. A. (2010). Perspectives in quality: designing the WHO surgical safety checklist. *International Journal for Quality in Health Care, 22*(5), 365–370.

West, S. G. (2009). Alternatives to randomized experiments. *Current Directions in Psychological Science, 18*(5), 299–304.

Williamson, T. (2013). Investigative interviewing and human rights in the war on terrorism. In T. Williamson (Ed.), *Investigative interviewing* (pp. 3–23). Routledge: New York.

Zedner, L. (2003). Too much security? *International Journal of the Sociology of Law, 31*(3), 155–184.

Brandon Langley is a PhD Candidate at the Institute of Criminology at the University of Cambridge. He is also a Superintendent with a large UK police force. His research interests are police legitimacy, procedural justice, and training.

Barak Ariel is a lecturer in Experimental Criminology at the Institute of Criminology at the University of Cambridge and an Associate Professor at the Institute of Criminology of the Hebrew University of Jerusalem. His research interests are experimental methods, policing and impact evaluation.

Justice Tankebe is a lecturer in Criminology at the Institute of Criminology, University of Cambridge, UK. His research interests include illegitimacy, procedural injustice, distributive injustice, corruption, and vigilantism.

Alex Sutherland is Chief Scientist and Director of Research and Evaluation at the Behavioural Insights Team.

Marcus Beale is an honorary Professor of Practice with the Keele Policing Academic Collaboration, Keele University, with research interests in crowd psychology and procedural justice. Previously he was an assistant chief constable with a large UK police force.

Roni Factor is an Associate Professor at the Institute of Criminology of the Hebrew University of Jerusalem. His research interests include the social mechanisms of high-risk and delinquent behaviors, traffic violations, and road traffic crashes, with particular attention to disparities across ethnic and racial groups. Other research interests focus on police-community relationships, the legitimacy of law enforcement institutions, and quantitative methodologies.

Cristobal Weinborn is Director of the Police Science Department at "Fundación Paz Ciudadana", a Chilean Foundation which aims to evaluate and implement public safety policies.

Affiliations

Brandon Langley [1] · **Barak Ariel** [1] · **Justice Tankebe** [1] · **Alex Sutherland** [3] · **Marcus Beale** [4] · **Roni Factor** [2] · **Cristobal Weinborn** [5]

Brandon Langley
brl29@cam.ac.uk

Justice Tankebe
JT340@cam.ac.uk

Alex Sutherland
alex.sutherland@bi.team

Marcus Beale
m.beale@keele.ac.uk

Roni Factor
rfactor@mail.huji.ac.il

Cristobal Weinborn
criswein@gmail.com

[1]   Institute of Criminology, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK

[2]   Faculty of Law, Hebrew University, Mount Scopus, 91905 Jerusalem, Israel

[3]   The Behavioral Insights Team, 4 Matthew Parker Street, London SW1H 9NP, UK

[4]   Keele Policing Academic Collaboration, Keele University, Keele, Newcastle ST5 5BG, UK

[5]   Institute of Criminology, Faculty of Law, Hebrew University, Mount Scopus, 91905 Jerusalem, Israel