Experiments in policing

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
European Journal of Policing Studies

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Experiments in policing: The challenge of context

Ben Bradford
Chris Giacomantonio
Sarah MacQueen

Introduction
Policing research has recently taken a significant ‘experimental turn’. Spilling out from an initial ‘niche’ association with particular tactics (e.g. hotspots policing, Weisburd et al. 2012), experimental methods have now a much wider set of applications, often focusing on crime reduction/prevention but also, more recently, police-community relations. In a field previously dominated by observational studies, experimental techniques hold great promise in terms of identifying the effect of policing interventions on crime and other variables. In a climate of apparently greater openness to research and ‘evidence-based’ policing, where substantial investment has been given to building the infrastructure of a police research economy, police organizations in the UK and beyond are increasingly committed to experimental evaluations of new policies and existing practices (as witnessed, for example, by the creation of the Society for Evidence Based Policing¹, which takes a strong pro-experimental stance).

Yet experiments, particularly those involving random assignment of the treatment, are not without attendant problems and challenges. Just some of the issues faced by researchers wishing to apply experimental methodologies listed by Shadish et al. (2002: 276) include: that certain types of questions (e.g. those involving historical events) cannot be addressed experimentally; randomization can be ethically challenging or impossible; experiments can often take a long time to produce answers; many important explanatory variables (e.g. age, gender) cannot be manipulated; and that experiments need to be built on extensive pre-existing bodies of work, and are thus often not well-suited to new and fast-changing situations. Many of these issues would seem particularly salient in policing contexts, marked as they often are by ethical dilemmas, a fast-changing practice and policy environment and the frequent need to produce rapid answers to emerging questions, often with already stretched resources.²

There is not inconsequentially a long history within academic criminology (and more widely in the social sciences) of resistance to, or at least skepticism about, experimental methods. This has stemmed from diverse sources: epistemological concerns about the ability of experimental methods to generate truly ‘accurate’ or indeed ‘useful’ knowledge (Hope 2009); debates about the ‘realism’ of experiments, i.e., about the extent to which the experimental method requires an abstraction from the normal course of events that is so extreme as to make subsequent interventions based on the resulting evidence inapplicable in other locations, times or circumstances (Pawson and Tilley 1997; c.f. Cartwright 2007); ethical issues relating to lack of attention to backfire and knock-on effects

¹ http://www.sebp.police.uk.
² Experimental studies in policing tend to be resource intensive due to a tendency for interventions to be designed to ‘up the dose’ of existing activities and practices or to require additional activities on top of normal duties (Weisburd 2000).
from policy interventions founded on experimental studies (Carr 2010, Kochel 2011); and concern that an over-emphasis on the ‘gold standard of experiments undermines the legacy of much seminal police research and on-going attempts to use more diverse methods (Lumsden and Goode 2016). Moreover, experimental and quasi-experimental prevention and intervention trials in criminology consistently exhibit preference toward the intervention when trials are led by intervention developers (Eisner 2009; cf. Sherman and Strang 2009), compared with trials led by independent researchers. This suggests that experimental data may be susceptible to experimenter bias, a point we return to later in the paper.

To be clear at the outset, we share many of these concerns. Experiments do not offer a panacea to criminological research. Rather, they should be considered as simply one item in the discipline’s methodological ‘tool-kit’ – and one that should be used with some care and caution. With this in mind, in this paper we consider some of the challenges to fielding experiments in policing, focusing in particular on questions of how, where, when and who. We present a detailed examination of two recent experimental studies of police-community relations, similar in focus and approach yet yielding very different sets of outcomes – the Scottish Community Engagement Trial (ScotCET), which considered the effect of procedurally just practice on public trust and police legitimacy (MacQueen and Bradford 2015, 2016), and ‘Making and Breaking Barriers’, which explored the effects of deploying mounted police in diverse contexts, including neighbourhood patrol (Giacomantonio et al. 2015).

As we describe below, both these studies involved close police-academic partnerships, and our main concern in this paper is broadly organizational. By building on the already published accounts of the two studies, and comparing and contrasting across them, we are able to generate insight not only into how they functioned as independent research projects but also into a much more general set of questions about how experimental research can ‘work’ in policing. We ask: what were the characteristics of the police organizations and partnerships involved that influenced the functioning and outcomes of the studies? How did the climate within which the police organizations were working affect the research? And where did the practice of individual police officers fit within these processes? Along the way we also touch on arguably more foundational issues concerning what ‘success’ means, and how to foster climates in which to meaningfully discuss the results of research, highlighting the importance of exploring and understanding the complexities of experimental studies and the mechanisms, relationships and processes underpinning their outcomes. We argue, in short, against a reductionist approach to the interpretation, and subsequent use, of the results from experimental studies (see Hough 2010; Sampson 2010).

The paper begins with an outline of the two studies, focusing on their core aims and outcomes and detailing similarities in development, method and context. We then move to consider how critical differences between ScotCET and ‘Making and Breaking Barriers’ may explain their divergent outcomes. We first address practical issues of design and delivery, examining the nature of the interventions alongside details of scale, location and experiment group composition to explore potential impacts on treatment maintenance and fidelity. We then consider important contextual issues, emphasizing in particular the timing of the studies in relation to fundamentally different organizational (and
political) climates. We also take the opportunity to reflect on the risks posed by research of this nature, which requires close police/academic partnership, and, therefore, which also underlines the necessity of critical distance between academics and practitioners. The paper concludes with consideration of lessons learned and reflections on the implications for the future of experimental (and other) policing research.

The two studies
The background to both studies covered in this paper was the same – the widely accepted view that public ‘trust and confidence’ in police, and the legitimacy the service commands, is important. On the one hand, there is much evidence to suggest that high levels of public trust and enhanced legitimacy feed into enhanced cooperation between police and public, greater compliance with the law and other desirable outcomes (Higginson and Mazerolle 2014; Tyler 2017). On the other hand, there are widespread concerns that public trust has declined over time and that, at least in some communities, police legitimacy has been eroded and even disappeared.³ In the UK there have over the last two decades been a slew of policies that have attempted to address this apparent disconnect, ranging from Neighbourhood and Reassurance Policing (Hughes and Rowe 2007; Innes 2004; Quinton and Morris 2008; Tuffin et al. 2006) to a more recent, tightened, focus on procedural justice as the prime factor motivating trust and legitimacy (Myhill and Quinton 2011; Stanko et al. 2012).

Both studies fit closely with this agenda, being concerned with promoting ways of ‘doing’ policing that might enhance public trust and police legitimacy. Moreover, both studies appeared to support an experimental approach, where a defined ‘treatment’ or intervention was to be applied within a specified population, and where the outcomes of that treatment could be measured. Although very different in terms of the nature of the intervention and its context, both were premised on the same underlying idea that the way policing is ‘delivered’ is important, and that changes in police practice will, under the right conditions, promote change in public perceptions of police.

The mounted police community patrol experiment
‘Making and Breaking Barriers’ (Giacomantonio et al. 2015) was the first ever in-depth academic study into mounted policing. Taking a multi-modal approach, the project explored the uses to which mounted police are put, their potential value across a variety of deployments, and the ways in which the public react to police horses in different environments. Of interest for current purposes was one part of the project that explored public responses to mounted community patrols (the modal deployment of police horses in England and Wales), with a particular emphasis on the extent to which this highly visible form of policing might influence ‘trust and confidence’. Police visibility is, at least in the UK, consistently associated with public trust – on average, people who perceive more visible policing express higher levels of trust (Bradford et al. 2009; Sindall and Sturgis 2013). Accordingly, a quasi-experiment investigating the effect of mounted community patrols on public trust in local areas was developed for the project.

³ Although overall public trust in police in England and Wales has increased significantly in recent years (ONS 2015).
The experiment proceeded as follows. Six areas, in three matched pairs, were selected for the study, four in Gloucestershire and two in south London. None of these areas had recent experience of mounted community patrols. One area in each pair (the test site) received a series of mounted patrols numbering seven or eight over a four-week period, while the control, area operated business as usual – i.e. the normal patrolling activities of police working in those areas carried on. Standard patrolling activity also continued in the test sites, therefore mounted patrols were an ‘addition’ to normal police activity.

Pre and post surveys of local residents living in the six areas demonstrated, first, that the introduction of mounted community patrols was noticed by local residents – in one test site 50 per cent of those surveyed in the post period were aware there had been recent mounted patrols in their area. Second, having recently seen mounted police was, overall, associated with significantly higher levels of trust and confidence. Third, public trust in police, across a range of indicators – effectiveness, community engagement, and overall confidence – increased in the test areas relative to the control areas after the mounted patrols had taken place, suggesting that this form of visible policing may indeed have a positive effect on trust and confidence.

A programme of Systematic Social Observation (SSO) was integrated into the MCPE. SSO was pioneered in policing studies by Reiss (1971) and later refined by Mastrofski and colleagues (1998) and Schulenberg (2012), but has rarely been used in the UK. A team of observers monitored officers on mounted and foot patrols during the experimental period, in both the test and control sites, to measure the quantity and quality of ‘engagement’ between members of the public and the police. An engagement was defined as any active noticing of, greeting, or interacting with police officers, and overall mounted police generated approximately six times as many such engagements compared to foot patrol officers over equivalent time periods; an average of 332 engagements per mounted officer per shift compared with 50 per foot officer per shift.

There are some important caveats to place around the results of the MCPE. First, the overall effect on public opinion came mainly from one test site in Gloucestershire, albeit that change in public opinions from pre- to post-periods was also observed elsewhere. Second, the ‘positive’ effect of the mounted patrols on public opinion was primarily the result of ‘buffering’; during the experimental period, trust and confidence in the police declined in the control sites but stayed constant or at most increased slightly in the test sites. It seems that the horseback patrols inhibited a fall in trust and confidence in the test sites that would have occurred had they not taken place; one interpretation being that they bolstered or shored up public trust and confidence in the police rather than necessarily increasing it. Third, observations of the mounted patrols during the experimental period revealed significant treatment migration in one of the test sites (in the city of Gloucester). Rather than staying to the area to which they had been assigned mounted officers working in this site patrolled the city centre, which will almost certainly have resulted in residents of the matched control site – also in Gloucester – witnessing and perhaps interacting with the mounted

---

4 Though subsequently used again in a UK study on police training by two of the authors (Giacomantonio et al 2016).
police (this was indeed picked up by the survey instrument). Most of the significant statistical effects outlined above only emerged once Gloucester was excluded from the analysis.

Another issue with the design of the MCPE was that it did not allow identification of the precise cause of the effects observed. Was it greater visibility that was coincidentally delivered by the mounted police, or was it specifically the use of horses as a way of delivering this visibility? The experiment could not determine whether, for example, a higher number of foot patrols might have had a similar effect on public trust. However, as noted the observations of foot and mounted community patrols conducted during the experiment suggested that the horses provided a specific form of visibility that generated substantially more engagement between members of the public and the police than ordinary foot patrols. The SSO data thus contextualized the findings of the experiment, and provided evidence suggesting it was visibility and the mode of deployment (on horseback, which many people seemed to like) that generated the buffering effect on public trust. By way of contrast, Ratcliffe et al. (2015) report that a hotspots intervention in Philadelphia had no discernable effect on satisfaction with police and perceptions of procedural justice in the treatment areas: increasing levels of non-mounted police in an area may have little effect on the opinions of residents, presumably because most people simply don’t notice them.

**ScotCET**

Unlike the MCPE, which was unique in terms of the mode of policing it evaluated, ScotCET was intended as a replication study. The Queensland Community Engagement Trial (QCET) was the first large-scale randomized field trial to explore the potential effect of procedurally fair policing during routine encounters with citizens. Results demonstrated that improving the quality of routine interactions between public and officers (during random breath testing operations) had a direct positive effect on public satisfaction with the process and outcome of the encounter, perceptions of police procedural fairness, overall trust and confidence in the police; and on police legitimacy (Mazerolle et al, 2012, 2013). The intervention comprised officers in the test condition delivering a scripted series of remarks, comments and questions designed to communicate the core elements of procedural justice, and QCET was one of the first studies to demonstrate a causal link between implementation of procedurally just forms of policing and change in public opinion.

In 2013 the Scottish Government funded ScotCET as, in part, an effort to replicate QCET. However it proved that direct replication was not possible, largely due to legislative and operational differences between Scottish and Australian policing (not least that random breath testing, without reasonable suspicion of drink-driving, is illegal in Scotland). Instead, an adapted design was developed in partnership with road police officers working across Scotland. The intention was to match QCET as closely as possible while developing a project that would ‘fit’ the Scottish context.

---

5 With a wider aim to inform the Government’s Justice Strategy for Scotland ‘Reassuring the Public’ programme.
The experiment involved all 20 road police units operating at the time, and took place during the four-week 2013-14 Festive Road Safety Campaign. This annual campaign is intended to cut drink driving and enhance road safety via a proactive approach to stopping drivers with defective vehicles or who are suspected of drink/drug driving. Units were randomly assigned to experiment or control conditions via a matched-pair design (with pairs defined largely via geographic proximity and/or similarity of operational focus). During the first week of the campaign all units conducted ‘business as usual’, simply distributing a questionnaire to drivers once the encounter had finished. A week long hiatus followed while materials were distributed to units in the experimental condition and officers were briefed on the intervention. They were asked to verbally deliver a set of key messages during encounters, and to distribute a leaflet designed to enhance perceptions of procedural justice (a leaflet was also included in the QCET intervention). The aim was to introduce a level of consistency to encounters such that each included all of the core elements of the procedural justice model, whilst allowing officers to continue acting responsively and naturally. Units in the control condition continued to operate ‘business as usual’, and all units continued to hand out questionnaires to the drivers they stopped.6

The overarching hypothesis was that the positive findings from the original QCET would be replicated. The intervention was expected to enhance perceptions of procedural justice and levels of satisfaction, trust, confidence and legitimacy in the experimental units compared with those in the control condition. In the event, findings were very different. In the control areas a consistent pattern of improvement in scores on key constructs over the course of the Festive Road Safety campaign emerged. Yet this pattern was not repeated in the experiment areas, such that relative to the control condition driver assessments of procedural justice during encounters and driver satisfaction fell in the experiment areas. The intervention therefore seemed to have diminished, not enhanced, drivers’ sense of procedural justice during their encounters with road police, and led to a relative decrease in driver satisfaction. Note that this effect was again a result of buffering, although in this case the intervention appears to have inhibited an uplift in public perceptions in the experimental areas, which was observed in the control areas and which would presumably have also occurred in the experiment areas had the intervention not been delivered.

Possible reasons for these apparently perverse results are discussed in MacQueen and Bradford (2016). In short, a follow up study concluded that, on the basis of group interviews with officers in the experiment units, communication breakdowns during the implementation phase led to misunderstandings about the aims and objectives of the experiment and about the requirements it placed on officers, ultimately leading to implementation failure. Within a context of reform and perceived organizational ‘injustice’, which we discuss in more detail below, officer non-compliance with the requirements of the intervention was moreover enabled by recourse to some

---

6 Although the response rate was very low (6.6% overall) there was no significant differences in the sample structure or in the response rate, pre to post or between experimental or control groups.
well-known aspects of ‘police culture’, most notably scepticism and cynicism about the motivation of senior officers, the knowledge and ability of external ‘experts’, and the possibility of more positive relations with the public. All in all it seems that the nature of the experimental intervention, and the time and place in which it was delivered, led not only to implementation failure but also, possibly, triggered a diffuse negative response from officers in the experimental units that shifted their attitudes and behaviours and resulted in the overall negative results outlined above.

**Comparing the two studies**
We have outlined above two studies that had similar aims, to enhance public trust and police legitimacy, and which employed broadly similar methodologies involving matched pairs and pre-post designs (albeit that one was a full RCT while the other did not randomize the treatment). Yet the extent to which each can be considered a ‘success’ is very different. In the case of the MCPE not only was the intervention successfully delivered, but results were in line with the research hypotheses. By contrast, in ScotCET delivery was unsuccessful, and moreover something about it seems to have triggered a negative effect on public opinion, contra the original hypotheses guiding the experiment. We now turn to an exploration of some of the potential reasons behind these different outcomes. What was it that enabled the MCPE to ‘work’ as intended, while ScotCET did not? Before we consider differences between the two studies that might explain this conundrum, however, it is important to outline some of the other factors they had in common.

**What (else) did ScotCET and the MCPE have in common?**
One important element linking the two experiments was that both drew on well-established theories and empirical research. In the case of the MCPE, this context was simple – the well-established empirical link between police visibility and public trust. ScotCET drew on the far more elaborate, although equally well supported, procedural justice model. For both studies, the “good deal of preliminary conceptual or empirical work” (Shadish et al. 2002: 277) needed for a successful experiment was in place.

A second similarity was the use of additional data to assist interpretation of the findings, SSO in the MCPE and group interviews in the case of ScotCET. Without this largely qualitative data, interpretation of the experimental data would have been less compelling, and wider readings of the headline findings difficult if not impossible. Recourse to additional data allowed the researchers to better identify the likely explanations for observed effects.

Another important factor, which as we discuss below might have been something of a mixed blessing, is that both projects benefitted from significant ‘buy in’ from senior officers. Experiments of this scale will simply not get off the ground unless they are backed by relevant decision-makers, but in both cases senior police partners invested significant time and resource into the project, and appeared committed to making it work.

Relatedly, both projects benefited from close, and cordial, relationships between academics and key police partners, both at senior ranks and at the more junior ranks tasked with implementing the projects. Similarly, in both cases police partners were closely involved in the design of the intervention, for
example via input into the list of key messages for the ScotCET intervention, or in helping select areas for the mounted police patrols. In both cases, the partnership was active rather than passive, and the final form of the experiment was the product of a joint programme of work rather than simply imposed by one side or the other. As in many other areas of policing and indeed criminological research, though, it is important to note that final decisions rested inevitably with the police. Moreover, since both studies required significant time and effort on the part of police, and willing cooperation from officers in terms of delivery, both could have been (and in the case of ScotCET almost certainly was) undermined or otherwise affected by actions or omissions taken at the street and/or managerial level. Indeed, while in the current cases this need to work ‘with’ police to deliver the research did not limit the research questions we were able to formulate, it is not hard to imagine cases where it might have done – an obvious example being experimental studies seeking to explore the potential negative effects of particular police tactics.

Finally, both experiments were fielded in a context of change. In Scotland this primarily related to organizational reform and, specifically, the formation of a single national force (Police Scotland) in 2013. At the time of ScotCET the effect of this reorganization was still working through the roads police units involved in the experiment. In the case of the MCPE the wider context was that of budget cuts in the climate of austerity imposed after the 2010 General Election. While all police in England and Wales were affected by austerity this was particularly stark in mounted units, many of which were explicitly or implicitly threatened with disbandment. We return to these points below, as the precise nature of the change facing police across the two experiments may have had important effects on the results they produced.

**What were the differences?**

Despite the similarities outlined above the two experiments were also, in many ways, very different. In this section we describe these differences, and suggest reasons why some, at least, may have influenced the outcomes of the two studies.

**Intervention design**

Firstly, the interventions, while plainly different in content, were also very different in form. The MCPE intervention was simple – officers did what they normally would have done during a community patrol. There were no new skills for them to learn, and little debate was needed on what the intervention should look like, how it should be delivered or the motivation behind it. Moreover, since none of the areas included in the study had recent experience of mounted community patrols, the intervention was also very different from business as usual.

The ScotCET intervention was more complex. While its reliance on verbal and written communication meant it also required no new skills development or learning on the part of the officers delivering it – although arguably it should have done – the intervention needed significant effort to develop via a process of deliberation between academic and police partners. The aim was to create a tool that was as close to QCET as possible whilst retaining local relevance. The ScotCET intervention asked officers to work all of a series of key messages into each and every encounter under the experimental condition, and therefore, at
least in theory, enforced significant change on their practice. Despite this, however, it seems in retrospect that the intervention was not that different from business as usual. Encounters between roads police and drivers in Scotland were already highly verbal (in contrast to QCET, where business as usual included very little interaction between citizen and police at all); roads police in Scotland were already quite good at ‘doing’ procedural justice, and the baseline survey showed driver satisfaction was usually very high. There may therefore have been a ceiling effect – it may have been difficult to shift already generally favourable views of drivers in a yet more positive direction. The ScotCET intervention therefore made significant demands on the officers delivering it, while at the same arguably failing to differentiate this ‘new’ practice sufficiently from what was already occurring.

Location and scale
There were further practical differences between the two studies. The size of the cohorts delivering them was very different – a small group of officers in the case of the MCPE (which involved approximately 20 mounted officers and a handful of foot patrol officers included in the observations) compared with, essentially, half of all roads police working in Scotland in the case of ScotCET (of whom there were around 240 in total at the time). The geographic spread of the two studies also differed hugely; tightly focused on local areas in the case of the MCPE and national in the case of ScotCET. Not only will these factors likely have led to far greater heterogeneity in the way the intervention was delivered in Scotland, but they also meant that basic processes such as briefing the officers involved were very much more difficult. This led among other things to a reliance on ‘cascade’ briefing, which seems not to have worked in several units (MacQueen and Bradford 2016): one reason for the implementation failure in ScotCET was that some officers in the experimental areas had simply not been told what to do, or had been given summarized instructions that failed to capture the relatively complex requirements of the experiment.

A related issue was the ability of the academic partners and indeed senior officers to monitor the activities of officers in the experimental conditions. The SSO conducted alongside the MCPE was vital in terms of understanding what happened, but also in identifying the treatment migration noted above. Yet the SSO was only possible because the times, locations, personnel and activities of the mounted patrols was known in advance. In the case of ScotCET the widely dispersed nature of roads traffic policing, the number of units and officers involved in the experiment, and indeed the sheer geographical reach of the study, meant it was impossible for the research team to monitor in any meaningful way what officers in the experimental units were doing. The often remarked fact that most policing is low visibility activity that occurs away from effective managerial oversight was underlined in ScotCET, and it is clear that many officers in the experimental group felt able to, and did, ignore the instructions they had been given (MacQueen and Bradford 2016).7

Delivering the intervention

7 By contrast an observer was present at all QCET stops in the experimental condition (Mazerolle 2012) to ensure treatment fidelity was maintained.
The general working practices of the officers involved in the two projects were also obviously different, and this may have affected their ability to deliver the interventions involved. Perhaps most importantly, roads police officers inevitably encounter members of the public most often in ‘confrontational’ circumstances, having pulled them over, stopped them or otherwise accused them of misbehavior, mistakes or carelessness. They are therefore used to handling encounters in a particular way in order to ‘manage’ such confrontation, and may have been reluctant to diverge from their standard practice (MacQueen and Bradford 2016). Mounted police officers, by contrast, are much more used to dealing with people in diverse non-enforcement, conversational situations. Indeed, this is a key aspect of the role, and there is likely to be an element of self-selection of officers into such a position, which requires a very public-facing, engagement heavy and ‘on display’ working role (among other things, they needed to be able to deal with the extremely high level of public attention they received in most of their deployments). As noted, it seems likely that the ability of the mounted officers to interact on a more informal basis with members of the public is one reason for the positive effects of the intervention.

Organizational climate
The importance of organizational climate in facilitating (and indeed sustaining) change in policing has been increasingly highlighted in recent research. The organizational environment in which police work, and particularly the ways in which officers are treated by the organizational hierarchy has been shown to influence their beliefs about how policing should be undertaken, and their willingness to adapt or change and enact key values, in particular those around the importance of community oriented and procedurally fair policing (e.g. Bradford and Quinton 2014; Haas et al 2015; Skogan et al 2015). Organizational climate is thus an important factor to consider when looking for explanations for success or failure in the development and implementation of experimental studies (Wain et al. 2017).

In ScotCET, organizational climate appears to have contributed to negativity in officers’ responses to the study and the requirements it placed upon them (MacQueen and Bradford 2016). In the face of minimal, inaccurate, or no communication about the aims of study and its requirements, officers ‘filled in the blanks’ themselves, drawing, it would seem, on their concurrent experience of a major organizational restructuring. Eight local police forces had recently been merged into a single national force (Police Scotland), led by a chief constable with a particular agenda for reform and unification of approaches to policing. Officers involved in the experiment reported changes in key objectives that were felt to be antithetical to local experience and wisdom, a climate of top down change and cuts, low support from senior management, and a sense of diminished morale. The negative reaction to ScotCET appears, in part, to have been due to a conflation of the study with these wider organizational processes. The experiment was seen as driven from the top of the organization, and to be the result of perceived failings on the part of operational officers and a need to impose new ways of ‘doing’ the job.

By contrast, the MPCE was as noted delivered by officers who felt mutually invested in the research enterprise. Recall that the context of austerity meant that many mounted police units were threatened with closure.
Collectively, this was a body of officers with a significant interest in demonstrating that mounted policing was useful, effective and an operational capacity that should be retained. This was certainly true of the senior ranks involved, but researchers found that such views were shared among Sergeants and PCs, including those involved in delivering the experiment. The officers delivering ScotCET had no such concerns, and were not motivated to demonstrate the worth of their job (at least as defined by the terms of the experiment) to external audiences or senior officers. In fact, ScotCET officers reported the imposition of competing priorities during the study period, describing a hardened, top-down led organizational emphasis on ‘stop and book, not speak’ running contra to the values and requirements embodied by the experiment as well as the operational wisdom of the officers. It is hard not to surmise that these differences in extrinsic and intrinsic motivations fed through into the way the two interventions were delivered in terms of officer commitment to, and consistency in delivering, their requirements.

The choice to experiment
The final difference between the two studies is arguably also the most foundational. Since both were concerned with the effect of (a change in) police behavior on public attitudes, an experimental method was, broadly speaking, well-suited to answering the research questions in both ScotCET and the MCPE. However, the wider set of factors affecting the choice of method were rather different. The MCPE arose ‘organically’ from the mounted police research project of which it was part. Both academic and police partners were interested in the effect mounted policing has on those who experience it, so the research questions came first and were then matched to the method. Its place within the wider project also meant that elements of the MCPE, particularly in the SSO, could be trialed and repeated in other contexts.

ScotCET, by contrast, was a stand-alone project motivated in the first instance by the Scottish Government. Here, the experiment in effect came first – the call for tender specifically stated that the project was to be a replication of QCET. Compared with the mounted project, therefore, the research team were limited in terms of what could be done, where, and with whom. MacQueen and Bradford (2015) underline that ScotCET was not, in the end, a replication of QCET, primarily because of the differences between policing in Australia and Scotland. Yet the programme specification limited the approach that could be taken: whatever happened, it would be an experiment, and it would match as closely as possible to QCET.

There may be an important lesson here. The limits placed around ScotCET could be an important reason why it did not work – perhaps it was simply impossible, or at least inevitably problematic, to transplant QCET from Australia to Scotland. An investigation into procedural justice policing in Scotland could have taken many other forms. Yet the policy context ensured that the experiment went ahead, with the results outlined above. Any drive toward the application of experimental methods as if this is automatically an appropriate way to proceed – rather than simply the best way to address particular research questions – may run into similar problems. The (methodological) cart should not be put before the (policy question) horse.
The research relationship: Considerations and cautions
The collaborative academic-practitioner relationships formed during these studies are common enough in experimental research, but differ from the relationships between researchers and police that characterize the wider sociological literature on policing. Traditional (generally qualitative) sociological studies of police have tended to be less concerned with technical evidence of ‘success’, and more with the relationships between police and groups within society and the forms of (in)justice evident in those relationships. Academics have often therefore had an ambiguous orientation toward the police. Yet experimental research by its nature accepts – at least to a degree – the notion of public policing as an inherent good, to be made better through evidence-based improvements. This does not, of course, mean that studies ‘for’ police – measuring the success of policing on its own terms\(^8\) – are inherently at odds with studies ‘of’ police (in Manning’s 2005; 2010 sense). Studies ‘for’ police do, however, raise issues around the intellectual independence of the researcher and the kinds of questions the research can address.

In both projects described here we worked extensively with police partners, who had input into the research questions and design (and indeed partly funded the MCPE). Our police partners were aware of the need for limits to their influence, and agreed at an early stage that results – favourable or otherwise – would be published soon after the completion of the research. Nonetheless we as researchers were also cognizant of our interest in helping police find ways to improve their practice. As Pawson and Tilley’s (1997) ‘realistic evaluation’ approach has highlighted, the kind of research we describe here deals with concrete problems and their potential solutions, a process that moves rather fluidly between measurement and recommendation. In both studies we wanted to ensure that the ‘treatment’ provided was given its best chance to succeed, and that our police colleagues were able to bring value to their organizations as a result of the research. While it was always the case that even if an intervention did not ‘work’ as expected police could use it to find ways to improve practice, we, as researchers, had something of a vested interest in the experiments ‘working’ as planned. In the spirit of reflexivity we should recognize that this impetus may have affected our reaction to the way the experiments ran and the results they produced.

Alongside measuring outcomes, both studies tried to derive usable lessons for police through examining the context in which they took place. The collection of additional qualitative data has not previously been central to the experimental paradigm (although see Sherman and Strang 2004), but was vital here. In this sense, both studies exhibited characteristics of process and outcome evaluation, and researchers’ input was used in both formative (pre-intervention) and summative (post-intervention) stages of the interventions. Yet, due to the collaborative orientation of these studies and all partners’ interest in facilitating incremental evidence-based improvement of policing practice, as originally implemented the research did not, and indeed could not, adopt a more critical orientation to the practices under examination. This is not because we would have rejected findings that were inherently critical of the police; it is because the research questions were not oriented to unearth such findings in the first place.

\(^8\) i.e. those of decision makers and managers that oversee police organizations.
In ScotCET a more critical stance to the police organization came with the follow up study, facilitating open engagement with the study context and fostering interrogation of its outcomes and issues. In the MCPE, it came not at all.

In turn, while we are comfortable in asserting that our research retained adequate intellectual distance between ourselves and the police, it would be naïve not to recognize and examine the nature of our relationships and consequent impacts on research results and reporting. As researchers, we enter into a complex ethical and political terrain when we work directly with police organizations. Collaborative relationships cannot be eschewed, however, simply because these kinds of studies are limited in their ability to generate fundamental critique of the various forms of modern policing.

Other fields where similar partnerships have emerged may provide notes of caution. In pharmaceutical and biomedical research, for example, partnerships between academic researchers and industry partners have been in place for decades, and their impacts on research findings have been well-documented. Meta-analysis has shown that industry-sponsored biomedical research is ‘of similar quality’ to independent research (Bekelman et al 2003). Industry-funded experimental studies, however, tend to show a preference for – i.e. present evidence in favour of – ‘new’ interventions over ‘traditional’, placebo, or status quo interventions at a much higher rate than studies funded by non-profits, academic grants, or other means (Lexchin et al 2003, Davidson 1986, Als-Nielsen et al 2003, Bekelman et al 2003).

It is unclear whether this bias in industry-funded research has resulted from biased interpretation of results (potentially stemming from conflict of interest) or from study design issues, where for example industry partners may be more likely to support studies where the novel treatment is ‘likely to succeed’. The imbalance of results may also stem from ‘publication bias’, whereby pharmaceutical companies withhold negative results more readily than independent researchers. Moreover the bias toward ‘new’ interventions in industry-sponsored research has been demonstrated in peer-reviewed and university-facilitated studies (Bekelman et al 2003), where presumably a minimum standard of research quality and rigour was achieved and peer review processes were in place; while important, peer review and good research design on their own appear to be insufficient to remove all forms of bias. The process by which questions and studies are designed (and who designs them), how data is shared, and how (especially negative) results are reported all shape the biases that appear in a body of literature.

The potential consequences of academic-practitioner research relationships should therefore be considered. Eisner (2009) identified systematic bias favouring new interventions in crime prevention initiatives where trials are ‘developer-led’, suggesting a possible conflict of interest in experimental and quasi-experimental designs where researchers are not independent. Conversely, Sherman and Strang (2009) argue that independent researchers may exhibit bias against new interventions, suggesting that the data on bias is equivocal. It remains nonetheless worthwhile to examine results for bias on a case-by-case basis. Some of the susceptibility to bias of pharmaceutical and biomedical industry research may stem from profit motivation, which may be less relevant (though not irrelevant) in policing and crime prevention (cf. Sherman and Strang 2009). Nonetheless, navigating the increasingly
collaborative policing research terrain is probably unavoidable, and in any case such research partnerships appear to us to be a good thing on the whole. But setting boundaries for police partners, ensuring independent publication of results and sharing of data wherever possible, and willingness among police decision-makers to accept, learn from and properly understand failed interventions will be as important as identifying 'what works'.

Conclusions
By comparing and contrasting our experiences of fielding two experiments in policing, we may glean a number of important lessons about not just the development and implementation of experimental studies but also the place of research and ‘science’ within the varied landscape of policing and police-academic partnerships. Some lessons relate to ‘delivering’ the experiment, while a second, arguably more foundational, group relate to questions concerning what fielding a successful experiment actually means and our collective ability to foster a climate within which to meaningfully discuss experiments and other forms of research.

Delivery
The implementation process in experiments is critical to treatment fidelity and the production of meaningful outcomes. ScotCET and the MCPE demonstrate that even in a receptive environment treatment fidelity in ‘real world’ policing is difficult to achieve and sustain. From the experiences faced, we identify four broad factors relating to the delivery of experimental interventions.

First and most obviously careful consideration needs to be given to the intervention itself. Is it like or unlike current practice? What is required from the police in order to deliver it, intellectually, practically, and financially? How will what officers are being asked to do – which will inevitably be ‘on top of’ or ‘instead of’ current practice – interact with their wider attitudes, ideas, concerns and practices (c.f. Lumsden and Goode 2016)?

Second, and consequently, there is a clear need to bring on board, motivate, and secure the commitment of the street level officers delivering the intervention. In the MCPE this was relatively easy to secure, and the experiment went smoothly. While every effort was made in ScotCET – via for example officer focus groups to aid with the design of the intervention, visits to the experimental units between pre and post periods – it subsequently became clear that sufficient commitment was not obtained, resulting in patchy implementation and a negative effect on officer motivation (MacQueen and Bradford 2016).

Such commitment will most readily be secured by adhering to principles of organizational justice – communicating clearly with the officers involved; allowing them maximal input into the design of the intervention, taking their concerns seriously and acting upon them – while at the same time bypassing as far as is possible the established chain of command. As Sherman (2009) argues a naïve top-down approach – where front-line officers are simply ordered to take part in an experiment – is unlikely to work, due not least to the widely identified disconnect between top level and operational policing and the resulting cynicism among front-line staff. Moreover, hierarchical police cultures with little scope for bottom up engagement appear hostile environments for the idea that effective research is best premised on continuous social interaction and communication.
A ‘parachute in, get out quick’ approach from researchers or senior police seems unlikely to generate sustained engagement from the frontline, nor likely to embed the underlying principle of research within the police organization as a whole.

Third, despite the need to involve and indeed empower the officers delivering the intervention it is also important that the researchers managing the experiment retain control over, or at least knowledge of, what transpires. Due to the discretionary power vested in police officers, the extremely varied nature of their work and its inherently low visibility, in order to ensure the fidelity of the experiment it is vital that the researchers know what was actually going on, not just what was meant to be going on or, indeed, what police claim they were doing. This was, again, relatively easy in the MCPE, both because the intervention was simple and localized and because the SSO was built into the project. In ScotCET such control was more or less absent – officers delivering the intervention were simply allowed to get on with at as they saw fit. There is a significant tension here that will require careful handling. On the one hand, rank-and-file police officers need to be convinced to willingly deliver experimental interventions. On the other hand, the nature of the experimental method and the reality of police work combine to ensure that they cannot simply be expected to get on with doing so. Some level of, and often close, oversight is required. This issue has recently been explored by Wain et al. (2017), who found that officers involved in delivering a hotspots intervention appear to have resented the loss of autonomy involved (i.e. being told to go to hotspots, and being GPS-tracked in order to ensure they did), resulting in less positive views of hotspot tactics and less favourable perceptions of organizational justice compared with officers in the control group (see also Wood et al. 2013). While control is a fundamental part of the experimental method, over- or miscalibrated-control might not only create practical problems but actually be antithetical to what the experiment is trying to achieve (e.g. convince police that hotspots policing is a useful and effective tactic).

Fourth, and in summary, experiments in policing cannot be separated from their context, whether of the unit or police organization involved or in a wider social and political sense. Not only will this be important in framing the experiment itself (why here? why now?), and in all likelihood the way it is funded, but it will also shape the way participants, police and public, react to the intervention. Most important from a practical viewpoint will be the way the strictures of the intervention are experienced and assessed by the officers tasked with delivering it. Consideration of context should trigger questions concerning whether this is the right time and place to do this piece of research, and whether there are competing demands and objectives that may interfere with successful implementation.

Wider concerns
Turning to the wider set of issues elicited by our discussion, there is first the important question of what ‘success’ means. Is an experiment successful when the intervention is successfully delivered (treatment fidelity maintained, no migration effects etc.), when it has a successfully identified effect, and/or when this effect supports the original hypotheses? Or can success in the wider endeavor of research come about in different ways? These questions are
particularly pertinent given the nature of the police-researcher relationship described above. Most experiments in policing are designed to test, and frequently effectively promote, a particular tactic or approach. Often, both parties have significant interests in the research hypotheses being supported, suggesting experiments that are successfully fielded but which throw up null or negative effects will be seen as ‘failures’ – and often therefore pushed to the back the queue for publication. We feel confident in asserting that as researchers we learnt as much or more from the ‘failure’ of ScotCET as we did from the ‘success’ of the MCPE; yet viewed from many perspectives the former was, indeed, a failure. If we had merely stopped at that a significant opportunity for understanding some of the wider issues around the experiment – and indeed within Police Scotland at the time – would have been missed. It should be clear, moreover, that the outcomes and implications of ScotCET and MPCE can only be adequately understood in the context of the interplay between who implemented the interventions, how (and where) the underlying theory was practically applied, and when the interventions were introduced within wider organizational contexts. We suspect that the same can be said of any experimental study in policing.

Future readings of such studies therefore need to avoid reducing interpretation of the results to the outcome measurements alone. Using ScotCET as a cautionary tale, at the time of writing a small number of academic papers have cited its findings. In all cases that we could identify, ScotCET features as an experiment testing the effect of procedurally just policing, which produced a negative and anomalous result unexpected within the wider field (e.g. Sahin et al 2016; Maguire et al 2016; Fildes et al 2017; Bates et al 2016). No further engagement is proffered. In the absence of the qualitative data exploring this outcome (the follow up study having only very recently been published) interpretation of the outcome of ScotCET and its implications has been hindered by an absence of guiding contextual explanation. This point also lends support to the case against (over) reliance on systematic reviews (Hough 2010) as sources of evidence: none of the qualitative elements of ScotCET or MCPE would be eligible for inclusion in such a review, rendering the experimental outcomes difficult to engage with and utilize.

Experimental studies of police interventions are thus founded on relationships between researchers, managers and front-line officers that bear on how implementation ultimately takes place, what the results might be and how these might be interpreted. These relationships are essential, not incidental, to the experiment, and accounting for them represents vital contextual data. Study design needs to attend to the complex and contingent social processes that allow the research to take place, and all involved in the research process will benefit from reflexive assessments of the impacts of those processes on the experimental results.

To conclude, while the experimental turn in policing research holds significant promise for future research and evaluation, ScotCET and the MCPE proffer something of a cautionary tale. To be sure, technical issues – notably, in the ScotCET intervention – are part of this story, but the wider and deeper message is that doing and interpreting experimental research in policing cannot be separated from wider organizational, economic and political contexts. A refusal to engage in experimental methods would be counter-productive for
academics seeking to influence policy and police managers trying to identify ‘what works’. But if such methods are not used with a clear-eyed view of the kinds of issues we have described above, implementation will likely be difficult and interpretation of results problematic: diminishing impact, at best, and producing actively misleading interpretations, at worst.

References


