What does it mean when we ask the public if they are "confident" in policing? The Trust, Fairness, Presence model of "public confidence"

Abstract

"Confidence" is widely taken to be a crucial measure of the relationship between citizens and public services such as policing. It is acknowledged confidence is multifaceted and hard to measure, but often discussions are based on one "headline" rating of a single item, for instance: "What is your level of confidence in..." The subsequent focus for research is explaining what might drive "confidence", or what it might predict. We are interested in a more fundamental question: what does it mean when we ask the public if they are "confident" in policing? To answer this, we analyse extensive and detailed survey data specifically designed to measure public confidence, within the jurisdiction of a UK police force - West Midlands Police. We develop then validate a three-part model of confidence as Trust, Fairness, Presence and find good evidence to support this model across different waves of the survey. This extends existing literature with implications for policy.

The question of public confidence is widely relevant because it takes us to the heart of the relationship between citizens and public institutions (Cook and Gronke, 2005; Miao, Newman, Schwarz and Xu, 2014). "Public confidence" has been used to describe attitudes in relation to: EU crisis management institutions (Persson, Parker and Widmalm, 2017); U.K. policies on business (Linsley, McMurray and Shrives, 2016), asylum (Boswell, 2015) and local government (James and Moseley, 2014); various branches of state government in the U.S. (Kelleher and Wolak, 2007); the Supreme Court (Caldeira, 1986) and lower courts (Benesh, 2006); Congress and the White House (Ladd and Bowman, 1998); social security (Cook, Jacobs, and Dukhong, 2010) and other U.S. institutions (Cook and Gronke, 2005). There are long standing concerns about declining confidence across institutions by the American public (Heise, 1985; Lipset, and Schneider, 1983), and these concerns are shared across Europe (Eriksson, 2012), and Canada and Japan, as shown in the World Values Surveys (see Pharr and Putnam, 2000). In these settings, confidence is often measured with a single item.[1]

"Public confidence" is a concern in the context of policing because the public are a key source of help and information. It is also important in any society that the public have confidence in the executive branch of government, because this underpins consent and

---

[1] For instance, from the General Social Survey (GSS) "I am going to name some institutions in this country. As far as the *people running* these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them?" (original emphasis) (Cook and Gronke, 2005).

cooperation (Tyler, 1990). There are overlaps in terms of what "public confidence" means across these different settings. Several considerations shown to be important in public institutions are also understood as key in policing: the role of personal experience (Cook et al., 2010), perceptions of independence (Benesh, 2006), effective communication (Bain, Robinson and Conser, 2014; Heise, 1985) and importance of procedural fairness and transparency (Tyler and Degoey, 1995). However, confidence in policing differs in important ways from confidence in other public institutions because of the nature of police work (Brown, 1988; Klockar, 1984; Mastrofski, 2004). With the power to use force (Davies and Thomas, 2008) and role in maintaining order (Bradford, Murphy and Jackson, 2014; Morrell and Currie, 2015; Stott, 2011), the police represent state authority in ways other public servants do not (Skogan and Frydl, 2004).

Given these unique powers, important contributions in understanding public confidence in policing emphasize procedural fairness (Tyler 1990; Tyler and Blader 2000; Tyler and Degoey, 1995; Tyler and Folger, 1980). Fair treatment involves affording: "respect, dignity, voice, equity and neutrality to people when they are interacting with authority figures" (Bradford, Hough and Jackson, 2018: 640). Whether treatment is deemed objectively fair or is subjectively perceived to be fair has been shown to be relationally complex because of socio-demographic factors such as gender (Brown and Heidensohn, 2000; Westmarland, 2002), ethnicity (Marks, 2008; Smith and Stephenson, 2005; Schuck and Rosenbaum, 2005) and age (Ren, Cao, Lovrich and Gaffney, 2005). Other factors, such as police culture (Myhill and Bradford, 2013), and location can influence perceptions of confidence and procedural fairness (Skogan, 2006; Skogan and Hartnett,

1997). These relational factors make it problematic to talk of "the public", there can be multiple publics (Callaghan and Wistow, 2006).[2]

Unlike other public servants at street-level (Lipsky, 2010), the police often act in public spaces rather than within institutions, so their work can be more visible (Morrell and Currie, 2015). They are a continual source of media interest. They also carry out many tasks that are readily understandable: patrolling, crowd management, directing traffic, ticketing, stops, searches, questioning, arrest, pursuit, emergency response. These features - visibility, media interest, work in public spaces, understandable tasks, powers and use of force, association with authority - mean the police are not "faceless" as public servants are sometimes portrayed (Bradford et al., 2018). On the contrary, policing is more readily personified than many institutions or kinds of work, in terms of the prototypical figure of the policeman who (as seen by some) maintains order and embodies moral authority (Sunshine and Tyler, 2003).

This personification is a form of stereotyping, for instance the tasks just listed (patrolling, ticketing etc.) invite a very simplistic picture of police work. Stereotypes can be hugely problematic, but they also provide a kind of conceptual scaffolding that can be the basis for a relationship (Hardin, 2006) and social identification (Sunshine and Tyler, 2003). When citizens encounter the police, this can have important effects on their sense of self

---

[2] See also Appendix 1 on differences in terms of gender, age, ethnicity in our data – each of which can be a basis for understanding multiple publics.

4

and identity – officers can make people feel they either belong or do not belong to society (Bradford et al., 2014).

As with literature in public administration more widely (Benesh, 2006; Cook, Jacobs, and Dukhong, 2010; Kelleher and Wolak, 2007), measurement of confidence is contested in work on policing, making it difficult to trace the effects of change or evaluate effectiveness (Bradford et al., 2014; Fitzgerald, 2010; Fleming and McLaughlin, 2012; Myhill and Quinton 2010; Stanko et al. 2012). This debate within policing connects with a number of long-standing questions about the nature of measures in public administration: should we use objective or subjective measures? are differences in public attitudes based on direct experience, or generalized beliefs, or stereotypes (Hinds, 2009)? are they at a neighbourhood, regional or national level? what is the relationship between related terms such as trust, confidence, procedural fairness, legitimacy?

Here we contribute to understanding by focusing on a fundamental theoretical question: the actual meaning of "public confidence" in policing, in search of a better measure. This contrasts with studies whose focus is a single-item measure of confidence, used either to identify predictors of that measure or to connect to outcome variables (Merry et al., 2012) – which has been the predominant focus not just for policing studies but for studies of other public institutions (see Benesh, 2006). We are able to contribute to this more fundamental theoretical question because we draw on data from a large-scale survey specifically designed to measure different aspects of public confidence in policing. The survey has been incrementally refined over 39 waves to incorporate a very large array of

relevant measures and potential proxies for confidence. In studying the final three waves of the survey, each administered to c.8,400 members of the public, we develop and test a three-part model of confidence incorporating dimensions of trust, fairness and presence. Below, we review the literature on public confidence in policing and outline the context and methodology for this survey, prior to presenting our analysis and contribution.

Public Confidence in Policing

In both the U.S. and UK, and in other settings, the closest synonym for public confidence in policing is trust (Barton and Beynon, 2015). Following Stanko et al. (2012: 321), initially we can differentiate between confidence: "an overall attitude or orientation that people have towards the police"; and trust: something "that people do - or do not". To trust the police is to hold beliefs about how officers will act in certain situations (Bradford, Jackson and Stanko, 2009), for instance that they will be fair and also effective (Jonathan-Zamir and Weisburd 2013; Kochel, Parks and Mastrofski, 2013; Tankebe, 2009, 2010). An expectation of fair treatment and also competence is consistent with a prevalent definition of trust across different disciplines - the willingness of a trustor to make themselves vulnerable to a trustee (PytlikZillig and Kiimbrough, 2015) which is rooted in expectations and evaluations of competency and benevolence.

This definition resonates with themes we introduced earlier - that policing as a public institution is more readily personified (Sunshine and Tyler, 2003), and that the police have unique powers over citizens (Davies and Thomas, 2008). On this view, then, to have confidence in the police will include expectations conducive to trust, but it is a

more general evaluation - of whether the job the police are doing is one that citizens can rely on. It connects with other definitions of confidence as an assessment of whether the police are doing a "good job" – more precisely the response to the survey item: "how good a job do people think the police do in their local area?" (Stanko and Bradford, 2009). This definition also links to widely used definitions of confidence as an assessment as to whether the police are proving effective in "reducing crime" and that accordingly "cops count" (Bratton and Malinowski, 2008: 259). More broadly still, because the police represent state power and authority (Skogan and Frydl, 2004), confidence in the police is likely to reflect broader social attitudes relating to moral consensus and social cohesion. Thus confidence could also be defined as the result of activity: "promoting identification by generating a positive sense of the individual's place in society" (Bradford et al., 2014: 528).

The context for this study, England, is a particularly revealing one in which to study public confidence, because there was a concerted policy focus on confidence in policing during the last years of the New Labour government: the "confidence agenda" (Lee and McGovern, 2013). This was most clearly expressed in the policy paper *From the Neighbourhood to the National: Policing Our Communities Together* (Home Office, 2008). At the heart of this paper was a commitment to abolish a host of measures of policing effectiveness and to "set only one top down national target for police forces – to deliver improved levels of public confidence" (p. 3). Because of the decentralization of English policing into different forces, there were differences across measures and methods used in these surveys. However, many forces used survey instruments that were

refined and developed over successive waves, incorporating insights from a growing literature focused specifically on confidence. These surveys connected to theoretical concepts such as trust, identification and procedural fairness (Bradford et al., 2009; Jackson and Sunshine, 2007; Tyler 2006; Tyler and Huo 2002), as well as operational concerns such as visibility.

Since 2010, the "confidence agenda" has faded given substantial reductions in budgets, which also led to a steep fall in officer numbers nationwide (Rogers, 2014; Sindall and Sturgis, 2013). It was formally abandoned as a national performance measure, though many forces and PCCs have retained levels of confidence as a target. We focus on a large and detailed dataset at West Midlands Police (WMP) - a police service serving 3 million people over a geographical area that combines metropolitan and rural areas, and areas of high and low household income. The next section provides background on this survey that we advocate as a better source of understanding "public confidence" than instruments we know of either in policing or in public administration more widely.

The West Midlands Police Confidence Surveys

Overview: WMP public confidence surveys were initially conducted at least every 6 months from 2004, when extensive consultation among various stakeholders was used to develop the first version. Interestingly, unlike many forces, and until very recently, WMP continued to invest in carrying out these surveys in the same depth and detail even after a change in administration, an abandoning of the national confidence target, and in the wake of cuts. This was, senior officers told us, because they found the cost was repaid in

greater understanding. Each wave has had a sample size of around 8,400 and the waves we analyse (2015-2016) are the culmination of a progressively sophisticated approach to attitudinal surveying, grounded in the relevant literature (e.g. Bradford et al., 2009). Over 39 waves, items have been retained, added or omitted according to whether they have proven useful in enhancing understanding of confidence. There are limitations with relying on pre-existing datasets, however it is important to emphasise that the items in the survey were informed by the relevant literature on confidence in policing and surveys of the Metropolitan Police Service (Stanko et al., 2012). As inspection of the surveys and interviews with the market research agency confirmed, there was what one might call an academic integrity to these survey instruments. For instance, where there was a battery of items making up an established scale, these items were repeated verbatim across different waves of the survey. Refinements were not just based on empirical outcomes but on a more rigorous experimental logic that preserved established insights from the academic literature (Stanko and Bradford, 2009).

Administration: The survey is carried out face to face, at people's homes using computer assisted personal interviewing (CAPI), by professional researchers trained in accordance with the (UK) Market Research Society's guidelines. Random checks ensure surveys are appropriately administered. WMP has 28 parliamentary constituencies, 300 surveys are administered in each constituency. These are aggregated to the level of neighbourhood policing units, each comprising at least two, and at most 4 constituencies. With such surveys representativeness is an important consideration. Quota sampling is used to achieve this and for the entire sample there is approximately a +/-1% sampling error in

9

terms of representativeness on the lines of ethnicity, age and gender. At the level of a neighbourhood policing unit (a smaller sample) this error is between 3% and 4%.

Content: The latest version of the survey is extremely extensive and wide-ranging, incorporating 61 separate sections designed to evaluate residents' views of police practices and tactics, and related measures. Questions evaluate people's perceptions or experiences of: types of crime, including anti-social behaviour (ASB); feelings of safety and worry about victimisation; perceptions of police responsiveness and reliability; ability of the police to deal with a range of crimes such as ASB, knife crime, gun crime, gangs; perceptions of police tactics like stop and search, overall perceptions of the force; perceived value for money; a word of mouth recommendation measure. Demographic measures include marital status, number of children, disability, gender, age, work status, type of work, ethnicity, religion. Overall the survey yields 370 variables. In the next section we show how we progressively develop a 3 dimensional model of confidence: Trust, Procedural fairness, Presence.

Analysis

As a first step, we separately and independently content analysed all the survey questions. This was to select questions whose meaning overlapped with our initial definition of public confidence (above), and drew on relevant literature we had identified across different public institutions (Cook and Gronke, 2005; Heise, 1985; Ladd and Bowman, 1998; Lipset, and Schneider, 1983; Pharr and Putnam, 2000), policing (Fitzgerald, 2010; Fleming and McLaughlin, 2012; Myhill and Quinton 2010; Stanko et al. 2012) and

related themes of trust (Bradford et al., 2018; PytlikZillig and Kiimbrough, 2015) and procedural fairness (Tyler 1990; Tyler and Degoey, 1995; Tyler and Folger, 1980).

We did not simply select items that we thought were likely to be correlated with confidence. This is important to emphasize since we were not looking at what we consider to be a second-order question: whether public confidence was associated with outcomes. Instead, we focused on the first-order, fundamental question of what public confidence means, making use of a uniquely detailed and extensive survey instrument to do so. Putting this another way, our first-order question is concerned with "criterion" validity, whereas work to date overwhelmingly looks at "predictor" validity (see Bozeman and Perrewé, 2001 for a fuller account of the difference between these forms of validity). Criterion validity - what this measure actually means - is a crucial question in public administration because many studies have relied on a single-item measure of public confidence to draw inferences about population level attitudes to public institutions (Benesh, 2006; Cook, Jacobs, and Dukhong, 2010; Kelleher and Wolak, 2007; see Lipset, and Schneider, 1983 for an earlier review).

In the underlying literature on psychometrics, the use of single-item (or "global") measures rather than multiple item scales ("sum of facet" measures) is widely understood to be problematic when the construct being measured is a complex one. Indeed, the reliance on single item measures in such circumstances has been described as a "fatal error" (Wanous, Reichers and Hudy, 1997: 247). Much of this literature has considered the relative merits of global versus sum of facet measures of job satisfaction (simply

because this is one of the most frequently measured constructs at work). Although there are arguments on both sides when it comes to measuring job satisfaction (with some favouring global measures, some favouring scales), "confidence" is a more complicated construct than job satisfaction because it is an assessment of a public institution rather than a reflection on direct experience. As Hoggett (2006: 175) explains, citizen's attitudes to public organizations are complex because the purpose of public organizations is contested, "public organizations have to contain much of what is disowned by the society [such as] value conflicts and moral ambivalence".

In common with other studies (Benesh, 2006; Cook et al., 2010; Kelleher and Wolak, 2007), the survey also incorporates a single item measure for public confidence: "Taking everything into account I have confidence in the police in this area" (on a 7-point Likert scale anchored by semantic opposites "completely agree" and "completely disagree" with semantic anchors on every point). We had a choice of including the single item measure of confidence in our pool of items or treating it as a separate variable. We chose to treat it as a separate variable, since this allows the possibility of both a global (single-item) measure of confidence and a sum-of-facet measure. It also allows a test of the reliability of the measures of confidence we developed.

We compared the results of our content analysis to arrive at a common pool of twenty-eight items (below). All are measured on a 7-point Likert scale with semantic opposites: satisfied... dissatisfied or agree... disagree. A word in each item is underlined because we used this as the variable name. Including the variable name helps to show what happens

to questions across the stages of analysis as the item pool was reduced in search of a more parsimonious measure.

1. *How satisfied are you...* With the level of foot <u>patrol</u> in your neighbourhood ; That the police in this area... are <u>friendly</u> and approachable ; make <u>fair</u> decisions when dealing with local problems ; Can be <u>relied</u> on to be there when you need them ; Would treat you with <u>respect</u> if you had contact with them for any reason ; Treat everyone fairly <u>regardless</u> of who they are ; Can be relied on to deal with <u>minor</u> crimes ; <u>Understand</u> the issues that affect this community ; Deal with things that <u>matter</u> to this community ; Work with people in the community to <u>improve</u> your neighbourhood.

2. *To what extent do you agree or disagree that West Midlands Police (WMP)...* provide good <u>value</u> for money ; is committed to <u>serving</u> the public ; understands what matters to <u>people</u> ; is open and <u>honest</u> ; is <u>modern</u> and forward looking ; has a positive <u>reputation</u> within the local community ; works hard to <u>engage</u> with local residents.

3. *To what extent do you agree or disagree that WMP's officers...* are committed to <u>serving</u> the public ; understand what matters to <u>people</u> ; are open and <u>honest</u> ; are <u>modern</u> and forward looking ; have a positive <u>reputation</u> within the local community ; work hard to <u>engage</u> with local residents.

4. *To what extent do you agree or disagree that...* You are <u>proud</u> to be served by WMP; to gain better engagement with communities WMP officers and staff should reflect the <u>diversity</u> of the community it serves ; WMP has good engagement with all sections of the community providing services that are <u>accessible</u> to all ; WMP delivers a service that demonstrates it understands the <u>diverse</u> needs of the community it serves ; WMP works hard to develop <u>cohesive</u> community responses to problems within your neighbourhood.

Reducing our pool of items to avoid duplication

We reduced our pool of items prior to the exploratory factor analysis to avoid some clear duplication in semantic content, which can cause problems in interpretation of factors (Dragovic, 2004). As our review indicates, public attitudes to policing can reflect specific experiences and encounters, or more generalized opinions (Bradford et al., 2014; Sunshine and Tyler, 2003). The survey was designed to differentiate between specific

encounters and general beliefs. But as a result there is some duplication. The items in paragraph 2 above refer to WMP i.e. the force generally and paragraph 3 uses six of the same items but refers to WMP's officers. Testing correlations among these near duplicate pairs indicates that values were much higher than all other correlations (with Pearson correlation coefficient (r) values ranging from 0.689 to 0.804) indicating potential redundancy. (Large samples have great advantages, but make "p" values uninformative, whereas "r"s stabilise as sample sizes increase (Lin, Lucas and Shmueli, 2013): for instance a correlation of only 0.041 is "highly significant" with this many respondents.)

Rather than use both sets of items, we included items referring to WMP not WMP officers (i.e. paragraph 2 above). Whilst it can be valuable to differentiate between "the police" as a whole and local police (Schuck and Rosenbaum, 2005), there are good reasons to look at force level when evaluating "public confidence". This level of analysis is appropriate for questions of governance in our setting, given the role of Police and Crime Commissioners (Sampson, 2012). Also, since many people do not come into contact with the police, the most widely applicable measures among "the public" will ask for opinions on the force as a whole, rather than officer behaviour. This choice also helps to differentiate more accurately between satisfaction (Miller and Davis, 2008; Skogan, 2005) - an outcome of actual encounters with police officers, and confidence as an overall attitude (Stanko et al., 2012).

One item, "WMP officers and staff should reflect the diversity of the community it serves" had very low correlations (all below 0.25) with the other items we selected in this

first stage of content analysis, and r = 0.158 with respect to our single item confidence measure. The underlying construct this is meant to tap, representativeness, is important in policy terms because police forces across the globe tend to have fewer people from minority ethnic communities (reference anonymized for peer review). This is sometimes proposed as an explanation for lower favourability ratings among some minority communities. Using ANOVA to compare means, we examined whether people from minority communities rated this item significantly more highly and did not find evidence for this. Speculatively, if it were worded more simply: "WMP should have more officers from minority backgrounds" the results may have differed across different ethnicities. On reflection this item seemed poorly worded and we rejected it from subsequent analysis.

Results of our Exploratory Factor Analysis

EFA is appropriate for assessing dimensionality and developing more parsimonious and meaningful multivariate measures from related items. There are many EFA methods available but larger sample sizes tend to reduce errors in choosing across these - if there is a clear structure (Hair et al., 2006). We include detail on the procedure in Appendix 1 rather than distract from the flow of the paper. The rotated component matrix in table 1 - overleaf for reviewers' convenience - shows the factors the items grouped into, as per Appendix 1 (this was after we had randomly split the sample).

Table 1 Three Factor Rotated Component Matrix (Orthogonal Rotation Varimax)

	Component		
	1	2	3
the level of foot patrol in your neighbourhood			.657
police are friendly and approachable			.835
police make fair decisions when dealing with local problems			.824
can be relied on to be there when you need them*	.642		
would treat you with respect if you had contact with them	.632		
treat everyone fairly regardless of who they are	.701		
can be relied on to deal with minor crimes	.672		
understand the issues that affect this community*	.726		
deal with things that matter to this community*	.736		
work with the community to improve your neighbourhood*	.664		
provide good value for money*	.578		
is committed to serving the public	.670		
understands what matters to people*	.702		
is open and honest	.716		
is modern and forward looking	.683		
has a positive reputation within the local community*	.687		
works hard to engage with local residents*	.659		
you are proud to be served by WMP	.526	.437	
engage with all sections of the community accessible to all		-.838	
demonstrates it understands the diverse needs of the community		-.859	
develops cohesive community responses to neighb'd problems		-.824	

*Cross-loads at threshold of 0.3

16

Above shows loadings above absolute 0.4. Some suggest a threshold of 0.32 (about 10% of variance) for interpreting factor analysis (Tabachnick and Fidell, 2007), but this is a question of judgement and the 0.4 cut-off provides a clearer illustration of three factor structure. Re-examining the one item that cross-loaded: "You are <u>proud</u> to be served by WMP", we felt it was not a good measure and removed this from further analysis. It is unclear why someone would be proud to be served by their police force - an emergency service provider for whom (in the UK) there is no alternative.

Further reducing our pool of items to improve parsimony using focal variables

There are no hard rules as to the optimal number of items for a factor (a minimum of three is recommended), but with 14 of 20 items loading on just one factor we wanted to see if it were possible to have a sharper and more specific measure of Component 1 above. This is preferable for parsimony and makes it easier for others to adopt these measures in other settings and thereby to test their replicability (Vogt, 2011). The alpha for these items was also very high at 0.938 indicating redundancy.

Reading across the questions, we could identify common features our review had highlighted as important in terms of trust, understood as willingness to make oneself (i) vulnerable (PytlikZillig and Kimbrough, 2015). Underpinning trust are beliefs that could be a basis for cooperation, including expectations about (ii) fairness (Tyler 1990; Tyler and Blader 2000; Tyler and Degoey, 1995) and competent (iii) performance (Jonathan-Zamir and Weisburd 2013; Tankebe, 2009, 2010). We identified eight focal variables that were especially salient to these three features. Focal variables can help to refine

scales where existing relevant theory indicates some questions in particular capture the essence of a factor (see DiStefano, Zhua and Mindrila, 2009). These focal variables were: respect ; regardless ; serving (which we felt were relevant across all three - vulnerability, fairness and performance); relied ; honest ; modern (each relevant to vulnerability and also performance); and minor ; value (relevant to performance).

As well as finding a theoretical basis for these items, we identified common features of "non-focal" items suggesting these might not prove to be precise measures. As asterisked inn table 1, most of these "non-focal" items cross-loaded at a 0.3 threshold – which is sometimes suggested as a guide in factor interpretation for model development (Field, 2009). Of more substantive theoretical importance, all but one of these had local markers of place ("community", "neighbourhood", "local community", "local residents"). Terms like neighbourhood are crucial operationally, and the survey is designed to guide respondents explicitly to consider at times local officers and at other times WMP as a whole. However, our analysis suggests distinctions between different markers of place are problematic - either because people do not differentiate between policing provision as a public service in this way, or because terms like "community" and "local" are open to interpretation.

To examine this further, we spoke to key staff from the agency responsible for administering these surveys, in particular separately interviewing two highly experienced researchers. Between them they had personally administered well over 1,000 surveys to the public. When asked to reflect on their experience, each independently volunteered

that interviewees struggled to interpret markers of place (like neighbourhood, local, community). We return to this in the discussion section. The one outstanding item "understand what matters to people" cross-loaded at 0.3 and we could not connect it as closely to themes of vulnerability, fairness or performance from the literature. Removing it did not threaten scale reliability as the alpha for these eight focal items was still high at 0.871. We used these eight focal items to measure component 1 in a confirmatory phase of testing (below).

Validation of Model through Confirmatory Factor Analysis

For the second stage of analysis we tested this model using Confirmatory Factor Analysis (CFA) (Hair et al., 2006; Schreiber et al., 2006). This was appropriate because the 3-factor model had good theoretical justification in terms of existing literature and the clear structure was also empirically compelling. We found good evidence validating our three-factor model, but for completeness also tested post-hoc modifications to this model, and two competing models. We then replicated this analysis with additional fieldwork - two further waves of the survey using larger samples (since these were not split into exploratory and validation sub-samples). Figure 1 (below for reviewers' convenience) shows results in graphic form, using data from the second half of our split sample. As with the EFA we detail more technical considerations including measures of model fit in Appendix 1.

Figure 1 A Three Component Measure of Confidence
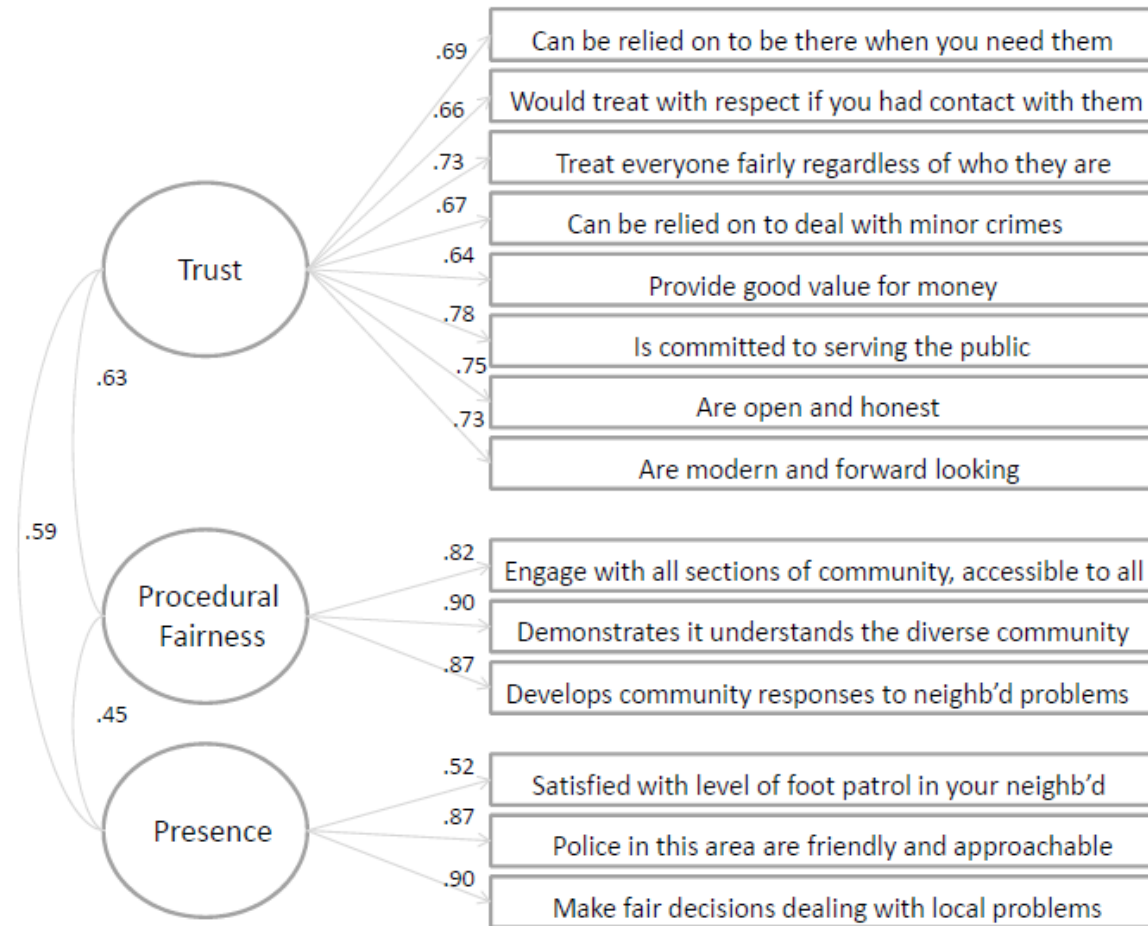
Alpha ratings for the subscales using the latest data are: Trust - 0.876; Procedural Fairness - 0.899; Presence - 0.732. We found these to be similarly reliable across gender, ethnicity and age (see Appendix 1). This is good evidence of convergent validity and suggests each could be the basis for measures in other settings, with necessary modifications to evaluate services other than policing.

It is worth pausing for a moment to reflect on the labelling and content of these three

factors. The first clearly relates to expectations and evaluations of competency and good intentions, and thus to the cognitive aspects of *trust*. The third equally clearly relates to a notion of place and *presence* that does not rely on potentially 'difficult' survey formulations relating to neighbourhood and community. The second, however, is slightly more difficult, as it refers to equity, engagement and collective working at, precisely, the community level. We have labelled this factor *procedural fairness* because it contains items referencing classic aspects of procedural justice, namely neutrality, lack of bias, and voice. But it is important to note that this is procedural justice at the community level (whatever respondents interpreted that to mean), not the personal – in this model, procedural justice at the individual level is mostly subsumed into trust. Although the items used are different, this factor may therefore relate closely to what Jackson and colleagues refer to as trust in police community engagement (e.g. Jackson et al. 2013). What we end up with, therefore, is a moderately clear distinction between individual level concerns – *trust* – and those that relate to (respondents' understandings of) neighbourhood concerns, *procedural fairness* and *presence*.

Finally, using mean scores for the items in each respective factor, we created three new variables. These explained fully 59% of the variance of our confidence item, and also 38% of the variance in a measure of word of mouth recommendation (I would say good things about WMP without being asked...). They also explained considerable amounts of the variance in measures we had initially selected in our content analysis and then dropped e.g. 44.3% of the variance in You are proud to be served by WMP (also see

Appendix 1).

Discussion

Our main theoretical contribution is to demonstrate the dimensionality of public confidence and in doing so propose a new way of measuring "public confidence", with subscales for each of three factors. This integrates and extends existing literature in policing and in public administration more widely. A multi-faceted scale supports the idea that confidence has several components and we differentiate more clearly between confidence and trust (Barton and Beynon, 2015; Fleming and McLaughlin, 2012; Myhill and Quinton 2010). This contributes to long standing debates about problems with interpreting expressed measures of citizen satisfaction in policing (Holdaway, 1982), and with other work in public administration on evaluating citizens' confidence in institutions (Cook and Gronke, 2005). It offers greater clarity about terms that are often used interchangeably in public administration (trust, confidence, satisfaction), but which we show in our setting can be defined more precisely and meaningfully.

To understand the relationship between citizens and public institutions better, our study shows the need to consider different components underpinning evaluations of "public confidence". We provide good evidence that confidence is multifaceted, so single-item measures will have limitations. We also show the need to consider whether attitudes to confidence are: generalized or specific; whether they are based on local experience or national issues; the extent to which a public service is actually understood; or whether it is personified and stereotyped; how perceptions of confidence relate to trust,

22

understandings of effectiveness and fairness; and the validity of subjective measures. These are all important insights into the meaning of "public confidence". Whilst we focus on developing a measure at the general level of "the public", the setting of policing shows assumptions need to be surfaced not just about the meaning of "public confidence" but about "the public". As our review indicates there are multiple publics who interact with services in different ways.

Predominantly, contributions to understanding "public confidence" highlight the importance of what we have called second-order questions - what actual ratings are, and how these can be tracked over time, or predicted, or related to other outcomes (Heise, 1985; Lipset, and Schneider, 1983). We show it is important to consider a fundamental, first-order question: what is happening when we ask "the public" to give ratings in response to specific items, but do not ask them to think about underlying concepts (which, we should stress, would be equally if not more problematic since it devolves the problems of definition to the survey respondent)? This setting, and our detailed measures, offers a basis from which to open up this question to other settings (Cook and Gronke, 2005) - for instance to evaluate perceptions of public institutions (Boswell, 2015; Linsley et al., 2016); different levels of government (James and Moseley, 2014; Kelleher and Wolak, 2007; Ladd and Bowman, 1998); the courts (Benesh, 2006; Caldeira, 1986), and European (Persson et al., 2017) or U.S. institutions (Cook, Jacobs, and Dukhong, 2010).

The three-factor solution offers clear support for the idea that generalized measures of appropriate intentions and competence (factor 1: Trust) are meaningful and also separable

from more specific questions to do with procedural fairness at the community level (factor 2: Procedural Fairness) or the local presence of police (factor 3: Presence). These findings essentially say that - at the level of the public as a whole - "public confidence" in "the police" is a combination of people's general beliefs about the police (Trust); their more specific evaluations of the fairness and appropriateness of police actions in their communities (Procedural Fairness); and localised accounts of what they directly experience (Presence). This multi-dimensional account of "public confidence" has different implications for understanding the relationship between the citizens and their police. We show this is not simply a question of teasing apart different aspects to confidence based on semantics. Instead, the beliefs that form confidence take shape at different spatial scales, from personal and local experience through to much more generalized attitudes.

Unlike a single-item measure, the three-factor model can be the basis for answers to an important strategic question: should police agencies try to improve public confidence by focusing on local tactics or improving the quality of police-citizen encounters, or should they try to influence wider factors associated with the media and social media by challenging stereotypes, establishing new stereotypes or emphasising learning points from failures? Whilst the easiest factor (in terms of tactics rather than cost) to influence may be presence (e.g. through reassurance patrols following incidents such as the attacks at London and Manchester), the multi-part structure to public confidence, and existing literature emphasise the need to consider procedural fairness and different drivers of trust.

An important consideration is at what level should we evaluate a "public service" when asking about "public confidence". Some institutions, like the Supreme Court (Caldeira, 1986) or the White House (Ladd and Bowman, 1998) are national, but in other cases, particularly in the absence of direct experience or contact - "placing" the institution can be more complex.

In the UK when evaluating public confidence in policing there are sound reasons for focusing on the level of the force (as we have here), but policing clearly takes shape at the local level, however defined. At the same time, our study suggests the public themselves can find it hard to differentiate between these levels of analysis. We found supplementary evidence that this was an ongoing challenge in research in the UK. In 2012/13 guidance to market researchers implementing the Crime Survey of England and Wales had included this question: "How long have you lived in this area? By this area, I mean within about a 15 minute walk from here." Interestingly, in later versions, guidance changes:

> I'd now like to ask you what you think about the police in this area.
> INTERVIEWER PROMPT IF NECESSARY: By local area I mean within a
> 15 minute walk from here (2015/16)

We find this detail revealing because the problem in defining "local" (or "community", "neighbourhood" or "this area") is either being evaded or devolved to the public. These terms are not self-evident but in this example researchers have left it to individuals to decide what they mean (see also Myhill et al. 2011). Our study illustrates there is a more general limitation - that when relying on single item measures, we do not really

know what people are thinking about when they answer them. This resonates with a long standing debate about the relationship between "subjective" and "objective" measures (Van de Walle and Van Ryzin, 2011). Any attitudinal survey (unlike some measures of crime and victimisation, say) always yields "subjective" measures. However, there is another layer to subjectivity because assumptions in how questions are framed can be problematic. The public is asked to base responses on differences that are objective from the perspective of the policy-maker: for instance to differentiate between local and regional policing. These distinctions are objective and operationally meaningful in terms of different policing tactics and priorities. But many people cannot distinguish between these or may interpret terms like "neighbourhood" differently. We feel this has wider relevance in understanding "the public". There is a further, perhaps an even more fundamental question, as to whether apparently objective terms are valid and whether surveys are really robust enough or sophisticated enough to account for multiple "publics".

In policing, just as in public administration more generally, "confidence" has been used in ways suggesting a multiplicity of meanings (Fitzgerald, 2010; Stanko et al. 2012). The underlying source for expressed confidence has been recognized as unclear (MacQueen and Bradford, 2015). Measures used to describe what counts as good policing (such as crime statistics, victim satisfaction, or public confidence) have tended to be single measures that are heavily contested and politicized (Loveday, 2006). These measures, to use a colloquial phrase from one of WMP's officers "take on a life of their own". In policing in the UK and elsewhere this can be problematic because it may mean

operational priorities are distorted. But here we have shown ample evidence of latent (not directly observable) dimensions that are multifaceted and stable and that tell us more about the underlying meaning of "confidence". Whereas these can be more difficult to use in determining operational priorities, they more faithfully represent the structure of the data and offer a comparatively parsimonious account of a complex setting and set of policy issues.

This contribution is timely in the UK given changes in the governance context and role of Police and Crime Commissioners (Lister, 2013; Rogers, 2013; Wood, in press). It also connects with more established literature in North America where debate about the political context for commissioners (typically not directly elected) is longer established (Hargrove and Glidewell, 1990; Sampson, 2012). In many models of policing, elections play an important part in police governance. Sometimes elections appoint those directly responsible for policing (for instance many U.S. Sheriffs). At other times people governing the police will be appointees of directly elected officials. Surveys of public confidence may be more informative about police governance than elections, where votes can be driven by many factors. This means we need greater understanding of the extent to which population level measures of public confidence are effective or can be improved. More broadly, if we value accountability of administrations and officials, it is critical to understand what is happening in large-scale surveys that ask "the public" if they have "confidence" in their institutions.

References

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family*, *67*(4), 1012-1028.

Bain, A., Robinson, B. K. and Conser, J. (2014). Perceptions of policing: Improving communication in local communities. *International Journal of Police Science & Management*, 16(4), 267-276.

Barton, H. and Beynon, M. J. (2015). Do the Citizens of Europe Trust Their Police? *International Journal of Emergency Services*, 4(1): 1-21.

Benesh, S. C. (2006). Understanding public confidence in American courts. *Journal of Politics*, *68*(3), 697-707.

Boswell, C. (2015). The double life of targets in public policy: disciplining and signalling in UK Asylum policy. *Public Administration*, *93*(2), 490-505.

Bradford, B. (2011). Convergence not divergence? Trends and trajectories in public contact and confidence in the police, *The British Journal of Criminology*, 51: 179-200.

Bradford, B., Jackson, J., and Hough, M. (2018). Trust in justice. in E Uslaner (ed.) *The Oxford Handbook of Social and Political Trust*. Oxford: Oxford University Press.

Bradford, B., Jackson, J. and Stanko, E.A. (2009). Contact and confidence: Revisiting the impact of public encounters with the police. *Policing and Society* 19: 20–46.

Bradford, B., Murphy, K., and Jackson, J. (2014). Officers as mirrors: Policing, procedural justice and the (re) production of social identity. *British Journal of Criminology*, *54*(4), 527-550.

Bratton, W. J. and Malinowski, S. W. (2008). Police performance management in

practice: Taking COMPSTAT to the next level. *Policing: A Journal of Policy and Practice*, 2(3), 259-265.

Britain, Great. (2012). Policing large scale disorder: lessons from the disturbances of August 2011, sixteenth report of session 2010-12, Vol. 2: Oral and written evidence. Stationery Office: London.

Brown, J. M. (2013). *The future of policing*. London: Routledge.

Brown, M. K. (1988). *Working the street: Police discretion and the dilemmas of reform*. New York: Russell Sage Foundation.

Brown, J. M. and Heidensohn, F. (2000). *Gender and policing*. Basingstoke: Macmillan.

Bryant, R., Bryant, S., Graca, S., O'Neill, M., Lawton-Barrett, K., Tong, S., Underwood, R. and Wood, D. (2012). (Eds). *Blackstone's Handbook for Policing Students*. 7e. Oxford: Oxford University Press.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Thousand Oaks, CA: Sage Publications.

Caldeira, G. A. (1986). Neither the purse nor the sword: Dynamics of public confidence in the Supreme Court. *American Political Science Review*, *80*(04), 1209-1226.

Callaghan, G. and Wistow, G. (2006). Publics, patients, citizens, consumers? Power and decision making in primary health care. *Public Administration*, *84*(3), 583-601.

Caless, B. (2011). *Policing at the Top. The roles, values and attitudes of chief police officers*. London: Policy Press.

Caless, B. and Tong, S. (2015). *Leading policing in Europe. An empirical study of strategic police leadership*. London: Policy Press.

Civitas (2015). *The Nine Principles of Policing*   available   on

http://www.civitas.org.uk/research/crime/facts-comments/principles-of-good-policing/ last accessed, July 18, 2017.

Cook, T. E., and Gronke, P. (2005). The skeptical American: revisiting the meanings of trust in government and confidence in institutions. *Journal of Politics*, 67(3): 784-803.

Cook, F. L., Jacobs, L. R., and Dukhong, K. (2010). Trusting What You Know: Information, Knowledge, and Confidence in Social Security. *The Journal of Politics* 72(2): 397-412.

Costello, A.B. and Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment*, Research and Evaluation, 10(7), 1-9.

Davies, A. and Thomas, R. (2008). Dixon of Dock Green got shot! Policing identity work and organizational change. *Public administration*, 86(3), 627-642.

Dillman, D., Eltinge, J. L., & Little, R. J. (Eds.). (2002). *Survey nonresponse*. New York: Wiley.

DiStefano, C., Zhu, M. and Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1-11.

Dragovic, M. (2004). Towards an improved measure of the Edinburgh Handedness Inventory: A one-factor congeneric measurement model using confirmatory factor analysis. *Laterality: Asymmetries of Body, Brain and Cognition*, 9(4), 411-419.

Drew, J. and Prenzler, T. (2015). *Contemporary Police Practice*, Oxford: Oxford University Press.

Eriksson, K. (2012). Self-Service Society: Participative Politics and New Forms Of

Governance. *Public Administration*, *90*(3), 685-698.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56-83.

Field, A. (2009). *Discovering Statistics Using SPSS: Introducing Statistical Method* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Fleming, J. and McLaughlin, E. (2012). Researching the confidence gap: Theory, method, policy. *Policing and Society*, 22(3), 261-269.

Floyd, F. J., and Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.

The Guardian (2012) *Reading the riots: Investigating England's summer of disorder*. http://www.guardian.co.uk/uk/series/reading-the-riots, last accessed, July 18, 2017.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (Vol. 6). Upper Saddle River, NJ: Pearson Prentice Hall.

Halliday, J. (2011). "London riots: BlackBerry to help police probe Messenger looting "role"." *Guardian Online*, http://www.guardian.co.uk/uk/2011/aug/08/london-riots-blackberry-messenger-looting.

Hargrove, E. C and Glidewell, J. C. (1990). *Impossible jobs in public management*. Lawrence, KS: University Press of Kansas.

Heise, J. A. (1985). Toward closing the confidence gap: An alternative approach to communication between public and government. *Public Administration Quarterly*, 196-217.

Hinds, L. (2009). Public satisfaction with police: The influence of general attitudes

and police-citizen encounters. *International Journal of Police Science & Management*, 11(1), 54-66.

Hoggett, P. (2006). Conflict, ambivalence, and the contested purpose of public organizations. *Human Relations*, 59(2), 175-194.

Holdaway, S. (1982). Police accountability: a current issue. *Public Administration*, *60*(1), 84-91.

Home Office (2008). *From the Neighbourhood to the National: Policing Our Communities Together*. London: Home Office.

Jackson, J. and Sunshine, J. (2007). Public Confidence in Policing: A Neo-Durkheimian Perspective, *British Journal of Criminology*, 47: 214-233.

Jackson, J., Bradford, B., Stanko, E.A. and Hohl, K. (2013). *Just Authority? Public Trust and Police Legitimacy*. Oxford: Routledge.

James, O. and Moseley, A. (2014). Does Performance Information About Public Services Affect Citizens' perceptions, Satisfaction, and Voice Behaviour? Field Experiments With Absolute And Relative Performance Information. *Public Administration*, *92*(2), 493-511.

Jonathan-Zamir, T. and Weisburd, D. (2013). The effects of security threats on antecedents of police legitimacy: Findings from a quasi-experiment in Israel. *Journal of Research in Crime and Delinquency*, *50*(1), 3-32.

Kelleher, C. A., & Wolak, J. (2007) Explaining public confidence in the branches of state government. *Political Research Quarterly* 60(4): 707-721.

Kim, J. O., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it*. Beverly Hills, CA: Sage.

Kline, P. (2002). *An easy guide to factor analysis*. London: Routledge.

Kochel, T. R., Parks, R. and Mastrofski, S. D. (2013). Examining police effectiveness as a precursor to legitimacy and cooperation with police. *Justice Quarterly*, *30*(5), 895-925.

Ladd, E. C. and Bowman, K. H. (1998). *What's Wrong; A Survey of American Satisfaction and Complaint*. Washington, D.C.: The AEI Press

Lee, M. and McGovern, A. (2013). Force to sell: Policing the image and manufacturing public confidence. *Policing and society*, *23*(2), 103-124.

Linsley, P., McMurray, R. and Shrives, P. (2016). Consultation in the policy process: Douglasian cultural theory and the development of accounting regulation in the face of crisis. *Public Administration*, *94*(4), 988-1004.

Lister, S. (2013). The new politics of the police: police and crime commissioners and the "operational independence" of the police. *Policing*, 7(3), 239-247.

Lipset, S. M. and Schneider, W. (1983). *The Confidence Gap*. New York: Free Press

Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public services.* 30th Anniversary expanded edition. New York: Russell Sage Foundation.

Loveday, B. (2006) "Policing performance: The Impact of Performance Measures and Targets on Police Forces in England and Wales", *International Journal of Police Science & Management*, 8 (4): 282-293.

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

Marks, M. (2008). Looking Different, Acting Different: Struggles for Equality within the South African Police Service. *Public Administration*, 86(3), 643-658.

Mastrofski, S. D. (2004). Controlling street-level police discretion. *Annals of the*

American Academy of Political and Social Science, 593, 100–18.

Merry, S., Power, N., McManus, M., & Alison, L. (2012). Drivers of public trust and confidence in police in the UK. *International journal of police science & management*, 14(2), 118-135.

Miao, Q., Newman, A., Schwarz, G. and Xu, L. (2014). Servant leadership, trust, and the organizational commitment of public sector employees in China. *Public Administration*, *92*(3), 727-743.

Miller, J. and Davis, R. C. (2008). Unpacking public attitudes to the police: Contrasting perceptions of misconduct with traditional measures of satisfaction. *International Journal of Police Science & Management*, 10(1), 9-22.

Ministry of Justice (2011). *Statistical bulletin on the public disorder of 6th to 9th August 2011*. London: The Ministry of Justice.

Ministry of Justice (2012). *Statistical bulletin on the public disorder of 6th to 9th August 2011-September 2012 update.* London: The Ministry of Justice.

Myhill, A. and Bradford, B. (2013). Overcoming cop culture? Organizational justice and police officers' attitudes toward the public. *Policing: An International Journal of Police Strategies & Management* 36: 338-56.

Myhill, A., Quinton, P., Bradford, B., Pool, A. and Sims, G. (2011). It depends what you mean by confident: Operationalizing measures of public confidence and the role of performance indicators, *Policing: A Journal of Policy and Practice* 5 (2): 114-124.

Persson, T., Parker, C. F. and Widmalm, S. (2017). Social trust, impartial administration and public confidence in EU crisis management institutions. *Public Administration*, *95*(1), 97-114.

Pharr, S. J., and Putnam, R. D. (Eds.). (2000). *Disaffected democracies: what's troubling the trilateral countries?*. Princeton University Press.

PytlikZillig, L. M. and Kimbrough, C. D. (2015). Consensus on Conceptualizations and Definitions of Trust: Are We There Yet?, in Shockley, E., Neal, T. M. S., PytlikZillig, L. and Bornstein, B. (eds.) *Interdisciplinary Perspectives on Trust: Towards Theoretical and Methodological Integration*. Springer, New York.

Reisig, M. D. and Kane, R. J. (Eds.). (2014). *The Oxford Handbook of Police and Policing*. Oxford University Press, USA.

Ren, L., Cao, L., Lovrich, N. and Gaffney, M. (2005). Linking confidence in the police with the performance of the police: Community policing can make a difference. *Journal of Criminal Justice*, *33*(1), 55-66.

Roberg, R., Novak, K., Cordner, G. and Smith, B. (2014). *Police &* Society, Oxford: Oxford University Press.

Rogers, C. (2014). Opinion: Police Accountability in the Age of Austerity, *The Police Journal: Theory, Practice and Principles*, 87(1): 1-2.

Rogers, C. and Thomas, G. (2013). *Applying the Peelian vision in the age of commissioners*, *Police Professional*, 349: 16-17.

Roycroft, M. (2013). A Blended Model For The Public–Private Provision Of Policing For England And Wales. In J.M. Brown (Ed) *The Future of Policing*, 191-202.

Sampson, F. (2012). Hail to the Chief? — How far does the Introduction of Elected Police Commissioners Herald a U.S.-Style Politicization of Policing for the U.K.? *Policing*, *6*(1), 4-15.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A. and King, J. (2006).

Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of educational research*, *99*(6), 323-338.

Schuck, A.M. and Rosenbaum, D.P. (2005), Global and neighbourhood attitudes toward the police: differentiation by race, ethnicity and type of contact, *Journal of Quantitative Criminology*, 21/4: 391-418.

Sentas, V. (2014). *Traces of terror: counter-terrorism law, policing, and race*. Oxford University Press.

Sindall, K. and Sturgis, P. (2013). Austerity policing: Is visibility more important than absolute numbers in determining public confidence in the police? *European Journal of Criminology*, 10(2): 137-153.

Skogan, W. and Frydl, K. (2004). Fairness and Effectiveness in Policing: The Evidence. *Washington DC: The National Academic Press*.

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

Skogan, W. G. and Hartnett, S. M. (1997). *Community policing, Chicago style*. Oxford University Press on Demand.

Smith, G. and Stephenson, S. (2005). The theory and practice of group representation: reflections on the governance of race equality in Birmingham. *Public Administration*, *83*(2), 323-343.

Stanko, E. A., and Bradford, B. (2009). Beyond measuring 'how good a job' police are doing: The MPS model of confidence in policing. *Policing: a journal of policy and practice*, 3(4), 322-330.

Stanko, B., Jackson, J., Bradford, B. and Hohl, K. (2012). A golden thread, a

presence amongst uniforms, and a good deal of data: studying public confidence in the London Metropolitan Police. *Policing and Society*, 22(3), 317-331.

Stenson, K and Silverstone, D (2013). Making police Accountable: governance and legitimacy, in J. M. Brown (ed.) *The Future of Policing*, Routledge: London: 429-445.

Stott, C. (2011). "Crowd Dynamics and Public Order Policing". In T. D. Madensen and J.

Sunshine, J. and Tyler, T. (2003). Moral solidarity, identification with the community, and the importance of procedural justice: The police as prototypical representatives of a group's moral values. *Social psychology quarterly*, 153-165.

Knutsson (Eds.), *Preventing crowd violence*. Boulder, CO: Lynne Reinner: 25-46.

Tankebe, J. (2009). Public Cooperation with the police in Ghana: Does procedural fairness matter? *Criminology* 47: 1265–93.

Tankebe, J. (2010). Public Confidence in the Police: Testing the Effects of Experience of Police Corruption in Ghana. *British Journal of Criminology* 50(2): 296-319

Tyler, T. R. 1990. *Why people obey the law*. New Haven, CT: Yale University Press.

Tyler, T. R. 2006. Psychological Perspectives on Legitimacy and Legitimation. *Annual Review of Psychology* 57: 375-400.

Tyler, T. R. and Blader, S. L. (2000). *Cooperation in groups: Procedural justice, social identity, and behavioral engagement*. Psychology Press: New York University.

Tyler, T. R, and Degoey, P. (1995). Collective restraint in social dilemmas: Procedural justice and social identification effects on support for authorities. *Journal of Personality and Social Psychology*, 69(3): 482-497.

Tyler, T. R. and Folger, R. (1980). Distributional and procedural aspects of satisfaction with citizen-police encounters. *Basic and Applied Social Psychology*, *1*(4), 281-292.

Van de Walle, S., & Van Ryzin, G. G. (2011). The Order of Questions in a Survey on Citizen Satisfaction with Public Services: Lessons From a Split-Ballot Experiment. *Public Administration*, *89*(4), 1436-1450.

Vogt, W. P. (2011). *SAGE Quantitative Research Methods*. Thousand Oaks, CA: SAGE

Westmarland, L. (2002). *Gender and policing*. London: Routledge.

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: how good are single-item measures?. *Journal of Applied Psychology*, 82(2), 247.

Wood D.A. (forthcoming) The importance of liberal values within policing: police and crime commissioners, police independence and the spectre of illiberal democracy, *Policing and Society: An International Journal of Research and Policy*.

Appendix 1: Exploratory and Confirmatory Factor Analysis, Demographics

Exploratory Factor Analysis (EFA): We randomly split the sample (wave 38) in two halves (of 4,250 cases) to allow separate exploratory and validation phases. Distributional assumptions (skew, kurtosis) were satisfied. Partly to enhance replicability, initially we used principal components analysis (PCA) with varimax and orthogonal rotation, by far the most popular method (Kim and Mueller, 1978; Kline, 2002). This method suggested a three factor solution when evaluated using: (a) variance explained (after rotation, the first factor explained 33.88% per cent of total variance, with subsequent factors explaining 15.53% and 12.64%); (b) cut-offs of 1 for Eigenvalues (does not always yield optimal solutions but was consistent with the three factor solution); and (c) the scree plot. Using conventional tests of suitability for EFA, Bartlett's sphericity test was highly significant, the KMO measure of sampling adequacy well above a suggested cut-off of 0.5, at 0.953. Assuming factors are uncorrelated (orthogonal rotation) is less realistic, and PCA is not always considered factor analysis as it does not discriminate between shared and unique variance of manifest variables (Costello and Osborne, 2005; Floyd and Widaman, 1995). However, these limitations tend to be exaggerated in studies with smaller samples. There is enough in common among various approaches that a clear structure is likely to be identifiable with any method used appropriately. As an additional step we used a variety of other methods of extraction and oblique rotation. The same three factor solution emerged. Table 1 shows grouping of variables (see Table 1).

Confirmatory Factor Analysis (CFA): After reducing the item pool (as explained in the paper), we tested the model from our EFA specifying a three factor structure. Having found support for this, we tested and rejected three competing models: a single factor structure, a two factor structure (where we allocated variables based on another EFA, specifying a 2 factor solution), and a version that addressed potential misspecification in error terms and dropped one low-loading variable (patrol). We used a range of common indices (excluding some chi-square based statistics which are vulnerable to sample size and would exaggerate fit). These indicated good fit for our 3-factor model, which used completely new scales and a very large, heterogeneous sample: GFI (0.931) and NFI (0.936) both above a suggested threshold of 0.9 (Byrne, 1994); CFI (0.94) just above a suggested threshold of 0.93 (Byrne, 1994), PCFI (0.750) below a suggested threshold of 0.8; RMSEA (0.079) just below 0.8 (Fan, Thompson and Wang, 1999). As a brief illustration, comparing the single factor / two factor models, fit indices were far worse, e.g.: GFI (0.738 single factor model / 0.800 two factor model), CFI (0.705 / 0.778). For a modified version we examined indices to look at covariances between error terms on the same factor and found some evidence of this in three paired terms: $e2 \Leftrightarrow e3$, $e7 \Leftrightarrow e8$, $e1 \Leftrightarrow e4$. Allowing these to covary and removing the comparatively low-loading patrol variable led to fit improvements: GFI (0.964), NFI (0.967), CFI (0.971), RMSEA (0.061) but at the expense of parsimony (PCFI worsened to 0.625). We judged these were modest improvements so retained the framework in Figure 1. We were cautious of making post hoc changes since this would undermine our claims to be validating a model generated in the EFA, and in effect be introducing a second, exploratory stage which is counter to the logic of CFA. To replicate our findings we carried out two further waves of fieldwork with larger samples (because we did not₃₉split these into exploratory and validation

phases). Fit indices improved, supporting the structure set out in figure 1: for wave 38 GFI was 0.971, for wave 39 GFI was 0.972. Also, in wave 39 we ran a second CFA model, this time imputing missing values (Acock, 2005; Groves, Dillman, Eltinge and Little, 2002) for "don't knows" to see if these were potential sources of systematic bias. These are not necessarily "missing data" as some answering will have had no police contact. Model structure was extremely similar (comparing path coefficients) as was model fit. (We did not include "don't knows" in correlation analyses, using pairwise deletion, but had over 6,000 responses for each bivariate correlation.)

Gender, Ethnicity, Age: *Gender* was binary, coded by interviewers (though for many, gender is neither binary nor identifiable from appearance). *Ethnicity* was self-chosen from 17 categories widely in use in social surveys in the UK and which we aggregated to 4 sub-groups used in the same schema: "White", "Mixed", "Asian", "Black" (else numbers in some categories were too small to test). Even after aggregating, and though the sample was representative, the "Mixed" (n=158, 1.9%) and "Black" groups (n=486, 5.8%) were much smaller in relation to "Whites" (n=6,175, 73.4%) and "Asians" (n=1,412, 16.8%) so we were cautious interpreting differences across categories (some respondents do not report ethnicity so percentages do not sum to 100). *Age* was self-reported ("What was your age on your last birthday?"). More procedures are possible with a ratio variable, we used a 65+/under 65 group to allow for easier comparison across gender, ethnicity, age. As mentioned "p" values are not helpful with such a large overall sample for interpretation, so we used mean scores in evaluating differences. Alphas for each sub-scale were high across *gender* (Presence 0.736 female, 0.727 male; Fairness 0.897 female, 0.901 male; and Trust both 0.876) *ethnicity* (Presence between 0.714 and 0.763; Fairness between 0.892 and 0.932; and Trust between 0.869 and 0.901) and *age* (Presence 0.723 65+, 0.730 under 65; Fairness 0.877 65+, 0.901 U65; and Trust 0.847 65+, 0.879 U65), evidence the measures were generalisable and stable. For *gender*, we found slightly higher ratings across the 3 factors among women (Trust - 5.5355 vs 5.4443 ; Fairness 5.2981 vs 5.2298 ; Presence 4.9666 vs 4.9182). That there were no stronger differences is in line with existing literature on the relationship between gender and attitudes to policing, where findings are mixed (see Jang, Joo and Zhao for a review). For *ethnicity* we found higher ratings for "Asians" on both our Trust factor (mean of 5.543 vs means of 5.5196 for "Whites", 5.2534 for "Mixeds", 5.2739 for "Blacks") and our Presence factor (mean of 5.0647 vs means of 4.9078 for "Whites", 4.6859 for "Mixeds", 5.0225 for "Blacks"). We found higher ratings for "Whites" on our Fairness factor (mean of 5.3268 vs means of 4.7470 for "Mixeds", 5.1586 for "Asians", 4.8670 for "Blacks"). Many studies find majority ethnic groups rate the police more favourably but it is outside the scope of the paper to make inferences about these differences or their generalisability to other settings, other than to note that it is common for much higher differences to be reported. For *age*, we found higher ratings among 65+ for Trust and Fairness but lower ratings for Presence (Trust - 5.5043 for under 65s vs 5.6396 over 65s; Fairness - 5.2545 for under 65s vs 5.5349 over 65s ; Presence 5.0232 for under 65s vs 4.9284 over 65s).