



Blurring the Distinction Between Empirical and Normative Legitimacy? A Methodological Commentary on ‘Police Legitimacy and Citizen Cooperation in China’

Jonathan Jackson¹  • Ben Bradford²

Received: 7 January 2019 / Revised: 12 April 2019 / Accepted: 9 May 2019 / Published online: 1 June 2019
© The Author(s) 2019

Abstract

In a fascinating study into the nature of police legitimacy in Southern China, Sun et al. (2018) present evidence that what researchers have previously been treated as possible sources of legitimacy—public perceptions of police conduct defined along the lines of procedural justice, distributive justice, effectiveness and lawfulness—are in fact constituent components of legitimacy. In this methodological commentary, we argue that the empirical strategy used to reach this conclusion is not fit for purpose because both conceptual stances—possible sources of legitimacy or constituent components of legitimacy—are consistent with the same fitted statistical model. Analysing nationally representative data from 30 countries across Europe and beyond, we also show that erroneous support for the approach to measurement is likely to be found wherever one looks. To be sensitive to cultural context means using a methodology that does not impose the preconditions of legitimacy, and we counsel against a trend starting in international criminology that does precisely the opposite.

Keywords Legitimacy · Police · Measurement · Procedural justice · Distributive justice

The ability of the police to assert social control and reproduce social order depends on the capacity to use force, so questions concerning the rightful use of this power are never far from the political surface. Normative concerns about police powers are ever-present in the debates that almost continuously roil around policing, yet police in liberal democracies rely on the legitimacy they command and the public cooperation, deference and compliance it engenders, and this raises *empirical* concerns about the extent to which the policed hold the police legitimate.

✉ Jonathan Jackson
j.p.jackson@lse.ac.uk

¹ Department of Methodology, LSE, Houghton Street, London WC2A 2AE, UK

² Department of Security and Crime Science, UCL, 35 Tavistock Square, London WC1H 9EZ, UK

The past decade and more has seen the pathbreaking US-based work of Tom Tyler and colleagues into the causes and consequences of public perceptions of police legitimacy spread across the world (Sunshine and Tyler 2003; Tyler 2006a, 2006b). Studies into the sources of police legitimacy have been carried out in social, political and legal contexts as diverse as Ghana, Finland, the Russian Federation, the UK, Pakistan, Sweden, Japan, Israel, Australia, Turkey, South Africa, France, Ukraine, China and Nigeria (for a review of the international literature see Jackson 2018). In many of these countries, legitimacy seems to rest a good deal on the extent to which police officers act in procedurally fair, neutral, transparent and trustworthy ways when making decisions and interacting with the public.

Sun et al.'s (2018) recently published paper on police legitimacy in a coastal city in Southern China is an excellent example of the increasingly international nature of this field of enquiry (for others see Tsushima and Hamai 2015; Kim et al. 2018; Akinlabi and Murphy 2018; Bradford and Jackson 2018; Gerber et al. 2018; and Oliveira et al. 2019). Sensitivity to context is important to their work. China is an authoritarian regime so it has 'low accountability and high coercion', hence people's feelings of obligation to obey external legal authority—often seen as a core aspect of empirical legitimacy (Tyler and Jackson 2013; Pösch et al. 2019)—may be complex and varied. But people may also judge the legitimacy of the police on the basis of procedural justice (fair process) *and* distributive justice (fair aggregate allocation of outcomes) *and* effectiveness *and* lawfulness:

...the police in an authoritarian state are commonly empowered with excessive authorities that do not match normative expectations of democratic policing (e.g., procedural fairness, institutional transparency, and accountability). Authoritarian policing is thus prone to abusive treatments of the public and state manipulative efforts of performance. Lawfulness, distributive justice, and effectiveness, originally proposed by Tyler as less imperative than procedural justice, could play a different or even an enlarged role in shaping police legitimacy under an authoritarian setting. (Sun et al. 2018: p. 2)

Drawing on data from a city-wide survey, Sun and colleagues conclude from their analysis that these four judgements are so strongly bound up with legitimacy that they collectively constitute the very construct itself. Rather than legitimacy being an overarching judgement about the right to power and the authority to govern—that may or may not be influenced by public judgements about whether police tend to act in procedurally just, distributively just, effective and lawful ways, as is conceptualised in procedural justice theory (Sunshine and Tyler 2003; Tyler 2006a, 2006b)—legitimacy *is* perceived procedural justice, perceived distributive justice, perceived effectiveness and perceived lawfulness.

Their paper raises important questions regarding the role that context plays in police–citizen relations, the conceptualisation and measurement of legitimacy, the difference between legitimacy and legitimation, and the position the researcher plays in (a) allowing the preconditions of legitimacy (the criteria that people use to judge legitimacy) to be an empirical question (i.e. discovered bottom-up) or (b) imposing the preconditions of legitimacy onto a given political community top-down. Their study also speaks to a broader issue of blurring the normative concept of legitimacy (the domain of the political philosopher, whose goal it is to consider from outside a given system what constitutes the ethical use of political power) with the empirical concept of legitimacy (the domain of the social scientist, whose goal it is to understand how people subject to power structures judge the rightfulness of authority).

Our goals in this paper are twofold. First, while we think that Sun et al.'s (2018) study makes a notable contribution to the literature, we question whether the analytical strategy they

use actually achieves what it is supposed to achieve. The statistical model they use is not a good adjudication tool because, we argue, the two conceptual stances under investigation in Sun et al. (2018)—(a) procedural justice, distributive justice, effectiveness and lawfulness are possible sources of legitimacy and (b) procedural justice, distributive justice, effectiveness and lawfulness are constituent components of legitimacy—are both consistent with the same fitted confirmatory factor analysis (CFA) model. We take the researcher through a thought experiment designed to illustrate the point that CFA is not fit for purpose because it cannot support the claim made by Sun and colleagues that ‘procedural justice variables should be considered as indicators of, rather than antecedents, of legitimacy’ (p. 14). We show that CFA is not a test one way or the other, for the simple reason that one would get exactly the same results *whatever one calls the collection of constructs*.

Second (and relatedly), we argue that their approach lacks cultural sensitivity because it is the outside experts, here Ivan Sun and colleagues, who are imposing the criteria that people use to judge institutional legitimacy. It is important not to conflate legitimacy and legitimation, we contend, and we counsel against a trend starting in criminology where researchers apply the approach of Sun et al. (2018) and Tankebe (2013) in new contexts, find that procedural justice, distributive justice, effectiveness and lawfulness are empirically distinct and positively correlated, and mistakenly interpret a successful fitted CFA model as evidence for the discovery that legitimacy *is* procedural justice, distributive justice, effectiveness and lawfulness. Indeed, our analysis of a 30-country dataset suggests that such an approach will be ‘successful’ wherever one looks. We find that in each country analysed the same fitted CFA model properties (as found in Sun et al. 2018) for comparable measures, yet this does not mean that legitimacy is so constituted in any particular context. It just means that the constructs are distinct and positively correlated, and that the measures have good scaling properties (it seems that we, as a community of criminologists, have good measures of these constructs and they are empirically distinct in most places we end up looking).

Researchers are, of course, free to impose onto a given context the criteria that people use to judge the legitimacy of the police. When one measures police legitimacy using survey indicators of procedural justice, distributive justice, effectiveness and lawfulness, one is doing just that—one is taking a normative position that procedural justice, distributive justice, effectiveness and lawfulness are so important that they *collectively constitute the judgement itself*. Our point is, however, that CFA does not do what Sun et al. (2018: p. 14) claim that it does. CFA cannot adjudicate between whether procedural justice, distributive justice, effectiveness and lawfulness are ‘indicators, rather than antecedents, of legitimacy’. Consequently, the approach of Sun et al. (2018) dictates the normative requirements of legitimacy under the smokescreen of (non-existent) empirical evidence.

The paper proceeds as follows. By way of conceptual ground-clearing, we distinguish between the normative concept of legitimacy of political philosophers and the empirical concept of legitimacy of social scientists. After summarising the standard approach to studying empirical police legitimacy, we review the original work of Tankebe (2013) who Sun and colleagues sought to replicate. We then discuss Sun et al.’s (2018) approach. Following the findings from our own empirical study into the measurement of legitimacy in 30 countries (that suggests that the measures of procedural justice, distributive justice, effectiveness and lawfulness scale reasonably well wherever one looks), we conclude with the idea that researchers who use CFA should be cognizant about what it can and cannot say about construct validity and, more broadly, that researchers should not unwittingly conflate legitimacy with potential legitimation.

Normative and Empirical Legitimacy

Political philosophers often employ legitimacy as a normatively laden term to describe whether state institutions meet an inherently value-based set of substantive criteria that specify how institutions ought to be configured and behave if their power is to be judged as rightfully held. In the case of the criminal justice system, a Western democratic conception of normative legitimacy might involve a group of outside experts coming together to decide that institutions should be judged according to principles of independence, transparency, accountability and other features of the rule of law. To take a measure of normative legitimacy, these experts might then collect national indicators of the rule of law.

By contrast, social scientists employ legitimacy as an empirically laden concept to describe whether—as a matter of fact—those that are subject to authority confer legitimacy on that authority (Tyler 2006a, 2006b; Caldeira and Gibson 1995; Gibson et al. 2003; Justice and Meares 2014; Meares 2017; Hamm et al. 2017; Trinkner et al. 2018). The empirical concept of legitimacy focuses on whether an institution finds ‘the approval of those who have to abide by it’ (Hinsch 2010: p. 40). Legitimacy is premised on a *fundamental accord* between rulers and ruled (Filiangeiri 1783–88, in Pardo 2000: p. 5) that is founded in shared norms and values and established via the ‘moral performance’ (Liebling 2004) of power-holders.

Social scientists typically operationalise police legitimacy along two connected lines:

1. Normative justifiability of power in the eyes of citizens (the right to rule): do citizens believe that the police as an institution is just, proper and appropriate?
2. Recognition of rightful authority (the authority to govern): do citizens believe that police officers are entitled to be obeyed?

On this account, empirical legitimacy is a process that involves acceptance (or rejection) of the implicit and explicit claims that police make to be legitimate. On the one hand, people judge the legitimacy of the police as an institution against the societal norms that dictate what appropriate conduct is (e.g. do police officers make neutral and objective decisions when dealing with citizens?). On the other hand, the content of legitimation (i.e. the bases on which legitimacy is justified or contested) is an empirical question—i.e. the content is not assumed by an outside expert on the basis of political, moral, legal, religious or some other philosophy. Indeed, what citizens of a particular social, political or legal context deem to be legitimising or deligitimising police conduct may vary from one country to another. For example, people in one context might judge the legitimacy of the police most keenly on the extent to which officers respect principles of fair process, while effectiveness might be more important in a different context.

The Standard Approach to Studying Empirical Legitimacy

When studying empirical legitimacy (what Tyler (2017), calls ‘popular legitimacy’), researchers (a) operationalise legitimacy as a psychological construct, (b) treat the normative appropriateness part of the legitimacy construct as an overarching judgement, and (c) allow the criteria that people use to judge legitimacy (i.e. type(s) of police conduct that shape people’s legitimacy beliefs) to be an empirical question (e.g. Sunshine and Tyler 2003; Tyler 2006a, 2006b). This strategy depends on distinguishing between potential sources of legitimacy (how officers are perceived to act) and overarching legitimacy judgements (whether the institution that these officers represent is deemed to have the right to power and authority to govern).

Because of the abstract nature of legitimacy, researchers tend to measure legitimacy by asking research participants whether authority is exercised appropriately (Tyler & Fagan, 2008; Jackson et al. 2013; Tyler and Jackson 2014). To avoid imposing the specific criteria that people use to judge appropriate police conduct, measures are worded generally in ways like ‘the police usually act in ways that are consistent with your sense of right and wrong’ and ‘the police generally have the same sense of right and wrong as I do’. Statistical analysis is then used to assess whether, for instance, procedural justice and effectiveness judgements are important predictors of legitimacy. If procedural justice judgements (do officers act in procedurally fair ways?) and effectiveness judgements (do officers act in effective ways?) explain significant variation in legitimacy, then the inference is that these are two *sources of legitimacy*, i.e. for authority to be appropriately exercised, then it needs to be employed in part in procedurally fair and effective ways.

In a recent example of this approach, Huq et al. (2017) analysed data from a nationally representative sample survey of the UK that measured the following constructs:

1. Public attitudes towards whether officers act in procedurally just ways;
2. Public attitudes towards whether officers act in distributively just ways;
3. Public attitudes towards whether officers act in effective ways;
4. Public attitudes towards whether officers respect the limits of their rightful authority;
5. Public attitudes towards whether officers use appropriate surveillance powers;
6. Perceived institutional legitimacy: normative alignment; and,
7. Perceived institutional legitimacy: duty to obey.

Huq and colleagues first used confirmatory factor analysis to assess whether the seven constructs (five being labelled as possible sources of legitimacy and two being labelled as constituent components of legitimacy, as specified by procedural justice theory) are empirically distinct albeit positively correlated constructs (this is an important first stage of analysis that we will return to later in the paper). Then having found evidence for empirical distinctiveness—and having found good scaling properties for the indicators of each construct—they used structural equation modelling to assess the extent to which each of the first five constructs (treated as possible sources of legitimacy) predicted normative alignment and duty to obey (treated as constituent components of legitimacy). Figure 1 provides an overview of the model.

Huq et al. (2017) found (a) that procedural justice and bounded authority were predictors of normative alignment and (b) that normative alignment and effectiveness were predictors of duty to obey. The conclusion they drew from this was that police legitimacy (operationalised as judgements of normative appropriateness) in the UK may be judged most strongly on the basis on which officers are seen to act in procedurally just ways and to respect the limits of their rightful authority. In other words, citizens seemed to judge institutional legitimacy of the police in part on whether officers respect the limits of their rightful authority, treat people with respect and dignity, talk and listen to people, and act in unbiased, transparent and accountable ways. In turn, normative alignment and police effectiveness were both predictors of willing consent to rightful authority, suggesting that duty to obey is linked to some degree with a popular sense of police strength in the fight against crime.

A majority of studies using this approach have found that the most important predictor of legitimacy is the extent to which people think that officers act in procedurally just ways—more important than for example perceptions of the effectiveness of the police in reducing crime and responding to victims. This is the case in the USA (Sunshine and Tyler 2003; Reisig et al.

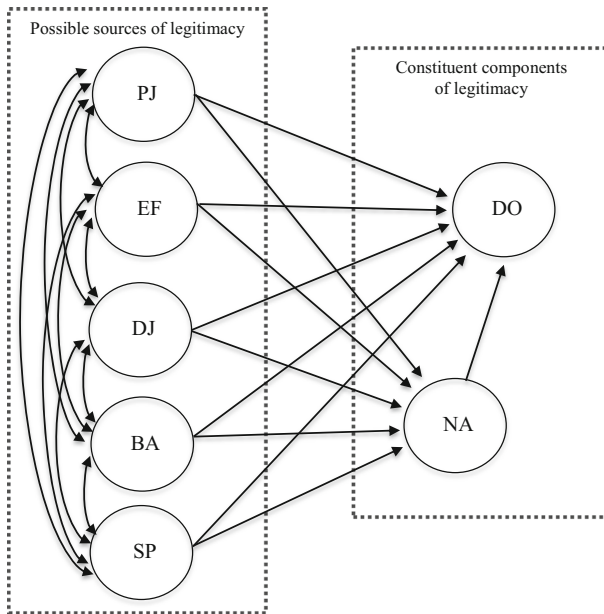


Fig. 1 Testing the possible sources of legitimacy. PJ=procedural justice; DJ=distributive justice; EFF=effectiveness; BA=bounded authority; SP=surveillance practices; NA=normative alignment; DO=duty to obey. CFA indicated that the seven constructs were empirically distinct and positively correlated, and that the various sets of multiple indicators had good scaling properties.

2007; White et al. 2016), UK (Huq et al. 2011; Bradford 2014), Australia (Murphy and Cherney 2012; Murphy et al. 2016, 2018), Israel (Mentovich et al. 2018), China (Sun et al. 2017) and Continental Europe (Hough et al. 2013a, 2013b; Dirikx and Van den Bulck 2013). Procedurally just police conduct seems to act as a legitimating process of justification, while procedurally unjust police conduct seems to act as a delegitimizing process of contestation (Tyler 2017). Notably, however, effectiveness and lawfulness judgements do seem to play a more important role in predicting empirical legitimacy in Pakistan and South Africa (Jackson et al. 2014; Bradford et al. 2014a, 2014b). In the latter cases, it may be that judgements about fair process can to some degree be crowded out by concerns about police (in)effectiveness and corruption, the sheer scale of the crime problem and/or the association of the police with a historically oppressive and underperforming state.

Police Legitimacy and Public Cooperation in Southern China

What, then, of Sun et al.'s (2018) study? Rather than using the approach outlined above, Sun and colleagues replicated Tankebe's (2013) empirical strategy, which sought to adjudicate between the following two conceptual stances:

1. Procedural justice, effectiveness, distributive justice and lawfulness are possible sources of legitimacy (where one then uses statistical modelling to determine the empirical importance of each one in explaining variation in the perceived right to power, see e.g. Huq et al. 2017); or,

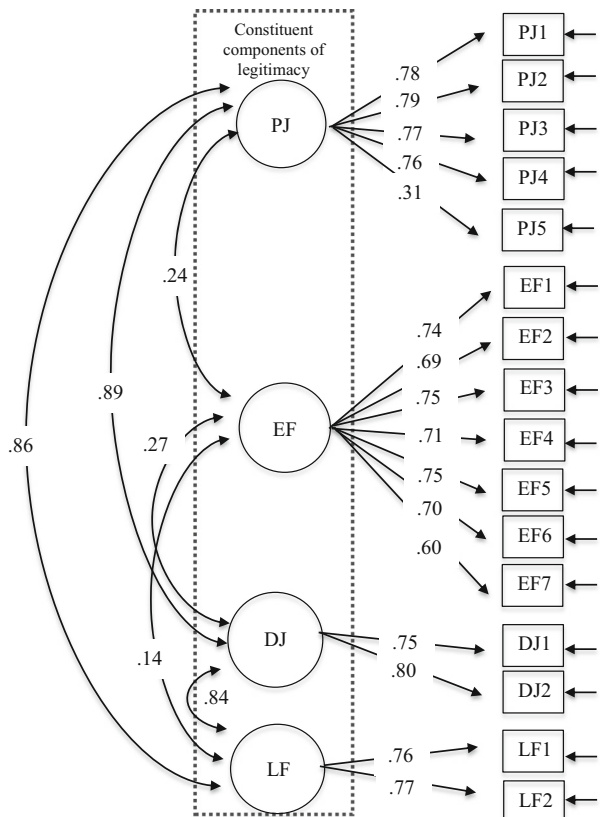
2. Procedural justice, effectiveness, distributive justice and lawfulness are constituent components of legitimacy (where they are so important that they collectively constitute the perceived right to power).

The details of the methodology used to adjudicate between these two positions are central to the current commentary. Because Sun et al. (2018) tried to replicate Tankebe’s (2013) London-based study, we review the original study first, before turning to Sun et al. (2018).

Tankebe (2013) drew upon data from a survey of Londoners (cf. Jackson et al. 2013a, 2013b) that measured among other things people’s perceptions of police procedural justice, distributive justice, effectiveness and lawfulness (using multiple indicators of each). He defined these as constituent components of legitimacy, and not possible sources of legitimacy. He found that a four-factor model that distinguished between the four constructs fitted the data reasonably well (summarised in Fig. 2); so, given the constraints in the fitted CFA model—i.e. no cross-loadings or error covariances and conditional independence of items given the four latent factors—one can treat these four judgments as distinct, albeit correlated, latent constructs.

Tankebe (2013) then tested whether effectiveness should be treated as a possible source of legitimacy or a constituent component of legitimacy. He fitted a three-factor CFA model without the effectiveness indicators and the effectiveness latent variable (i.e. he dropped EF1–EF7 and the EF latent variable in Fig. 2). He showed that the three-factor model fitted the data

Fig. 2 Tankebe’s (2013) measurement model of legitimacy. Source: Tankebe (2013). Viewing things differently: The dimensions of public perceptions of legitimacy. *Criminology*, 51, pp. 103–135. PJ = procedural justice; EF = effectiveness; DJ = distributive justice; LF = lawfulness. Fit statistics: $\chi^2 = 1056$, $df = 98$, $p < .005$; $CFI = .970$; $RMSEA = .044$ (95% CI .041–.046). The model fitted the data reasonably well and the measures had good scaling properties. For Tankebe (2013), this constitutes evidence that what is being measured here must be ‘constituent components of legitimacy’, and not ‘possible sources of legitimacy’



reasonably well. He ran a chi-square difference test to compare the relative fit of the three-factor and four-factor models. Finding that both the three-factor model (when indicators of procedural fairness, distributive fairness and lawfulness were included) and the four-factor model (when indicators of effectiveness, procedural fairness, distributive fairness and lawfulness were included) both fitted the data reasonably well, he deduced that:

Effectiveness has to be viewed as a component of legitimacy; police organizations that seek legitimacy must demonstrate effectiveness as a normative requirement. Coicaud ... has put this well: “Every political ruler who seeks to prove he possesses the right to govern [that is, is legitimate] has to satisfy, to try to satisfy, or to pretend to satisfy the needs of the members of the community.” For the police, those needs include safety and security. (Tankebe 2013, p. 121)

Tankebe’s (2013) reasoning was as follows. If the four-factor CFA model (Fig. 1) fitted the data reasonably well, then these four constructs should be labelled constituent components of legitimacy and not possible sources of legitimacy. The fact that the items scaled well, the fact that the analysis supported the idea that there were four underlying dimensions to the data, and the fact that these four factors were strongly and positively correlated (and that including effectiveness did not decrease the fit of the model) all meant that legitimacy *is* procedural justice, distributive justice, effectiveness and lawfulness.

Overall, the findings suggest that what police researchers have persistently tended to use as predictors of legitimacy (procedural fairness, distributive fairness, lawfulness, and effectiveness) are rather the constituent parts of legitimacy ... The results of the confirmatory factor analysis presented in this study suggest that the debate [about whether legitimacy causes procedural justice or procedural justice causes legitimacy] might be redundant because procedural fairness is a constituent part of legitimacy rather than something apart from it. (Tankebe 2013, p. 125)

In a replication study, Sun et al. (2018) motivated their empirical strategy by claiming that the approach of Tankebe (2013) offers a greater level of cultural sensitivity than the traditional approach to measuring and modelling legitimacy (e.g. Huq et al. 2017):

Drawing upon the dialog conception of legitimacy which views legitimacy “as always dialogic and relational in character” (Bottoms and Tankebe 2012, p. 129), Tankebe emphasized the dynamic nature embedded in police-citizen encounters, arguing that the different dimensions of legitimacy tend to have different effects across societies and among social groups within the same society. It is with this same embracement of cultural diversity that we attempt to test Tankebe’s work in the Chinese context. (Sun et al. 2018: p. 5)

To test the idea that the legitimation of the police is more complex in China than it is in contexts like the US, UK and Australia (where procedural justice is the key predictor of legitimacy), they examined whether Tankebe’s (2013) approach to measurement ‘worked’ in this city in Southern China. Like Tankebe (2013), they a priori defined legitimacy as procedural justice, distributive justice, effectiveness and lawfulness. Like Tankebe (2013), they used confirmatory factor analysis to test a four-factor model, although this time they included a second-order factor that they labelled *legitimacy*. This meant testing the idea that the second-order factor explained the correlations between procedural justice, distributive justice, effectiveness and lawfulness (specifically that the bivariate correlations can be modelled according to one underlying latent construct). Of note is the fact that Sun and colleagues could just as reasonably tested a model with two or three second-order

factors, since there is no requirement in the literature for legitimacy to be unidimensional (for discussion, see Jackson and Gau 2015).

Sun et al. (2018) found that the model (reproduced in Fig. 3) fitted the data reasonably well, with (i) good scaling properties for each of the four sets of indicators, (ii) four factors that were strongly and positively correlated and (iii) a second-order factor linked to each of the four first-order factors. Following the line of reasoning of Tankebe (2013) (see also Tankebe et al. (2016)), they interpreted the findings as follows:

...the convergent validity, discriminant validity, and internal consistency of all key measures were supported in the CFA analysis and the reliability tests...Substantively, these results mean two things: (1) procedural justice, distributive justice, effectiveness, and lawfulness are four distinct sub-constructs of legitimacy, and each sub-construct is well explained by its own corresponding observed variables, rather than by variables from a different sub-construct, and (2) the four sub-constructs correlate well with one another within their latent construct legitimacy. In short, Tankebe’s argument that procedural justice variables should be considered as indicators, rather than antecedents, of legitimacy, is supported. (Sun et al. 2018: p. 14).

It seems, then, that we have empirical evidence on what police legitimacy *is* (at least in the context of London and this unnamed city in Southern China). Legitimacy is not some overarching

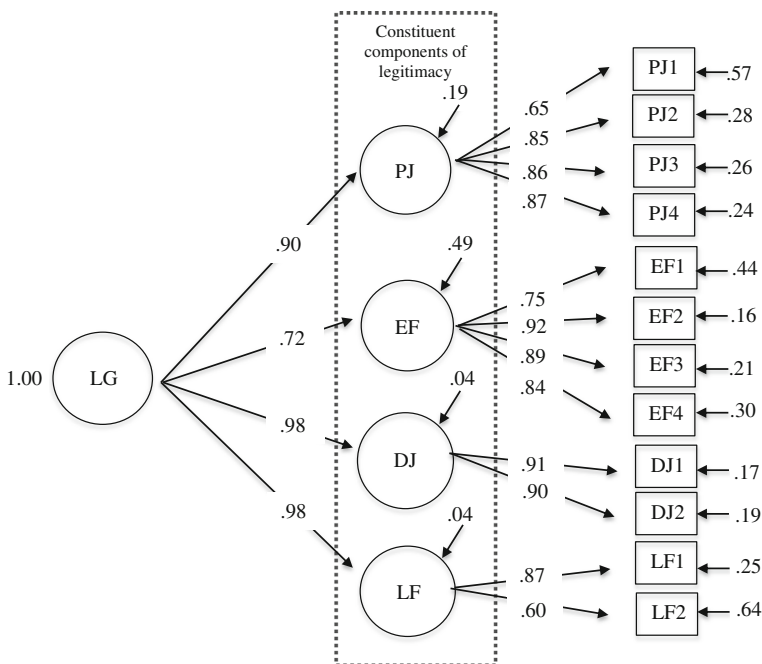


Fig. 3 Sun et al.’s (2018) measurement model of legitimacy. Source: Sun et al. (2018). Police legitimacy and citizen cooperation in China: testing an alternative model. *Asian Journal of Criminology*, 13(4), pp. 275–291. LG = legitimacy; PJ = procedural justice; EF = effectiveness; DJ = distributive justice; LF = lawfulness. Fit statistics: $\chi^2 = 206$, $df = 50$, $p < .005$; CFI = .982; TLI = .976; RMSEA = .058; and SRMR = .030. Because this model fitted the data reasonably well and because the measures had good scaling properties, Sun et al. (2018) argued that the second-order factor must be labelled ‘constituent components of legitimacy’ and not ‘possible sources of legitimacy’

judgement about the normative appropriateness of the police, coupled with a felt moral duty to obey legal authorities, as specified by Sunshine and Tyler (2003), Tyler (2006a, 2006b), Huq et al. (2017) and others. Legitimacy may not be *predicted* more strongly by procedural justice than by distributive justice, effectiveness and lawfulness, as is typically found in extant work. Strikingly, the CFA modelling in Tankebe (2013) and Sun et al. (2018) seems to show that legitimacy *is* procedural justice, distributive justice, effectiveness and lawfulness. It follows that if the police are to be judged to be legitimate in the current two contexts (see also Tankebe et al.'s 2016 work in the US and Ghana), then police officers need to be seen by citizens to act in ways that are (a) procedurally just, (b) distributively just, (c) effective and (d) lawful (to, one assumes, a roughly equal extent). In other words, police should prioritise visibly acting in procedurally just, distributively just, effective and lawful ways if they are to be seen as legitimate by those they protect and serve.

Normative or Empirical Concepts of Legitimacy?

Their paper was so thought-provoking that we asked Sun and colleagues if they could share the data. Given the fundamental importance of the finding regarding measurement, we were especially interested in the CFA modelling. Yet, while the dataset was not forthcoming, it turned out that we did not need it—and this only underlines the point we would like to make.

To explain, imagine you are embarking on a new study into police legitimacy in a coastal city in South China. You begin with the received wisdom on the nature of legitimacy. Sun et al. (2018) established that the residents of this coastal Chinese city judge the legitimacy of the police on the (roughly equal) bases of procedural justice, distributive justice, effectiveness and lawfulness. This key piece of work treated legitimacy as the joint distribution of these four constituent parts (i.e. legitimacy was represented as a second-order latent variable); legitimacy predicted cooperation; and the statistical effect was partly mediated by obligation to obey, which was treated as a potential outcome rather than constituent part of legitimacy (note that this second part of the study does not concern us here).

You are interested in questioning the status quo established by Sun et al. (2018). You want to test a measurement model that (a) operationalises legitimacy as a general belief that the police are morally entitled to dictate appropriate behaviour (i.e. as obligation to obey, see Tyler and Jackson 2013), (b) treats procedural justice, distributive justice, effectiveness and lawfulness as possible sources of legitimacy, which means (c) that it is an empirical question whether citizens of this coastal Chinese city judge the right of the police to dictate appropriate behaviour (i.e. obligation to obey) on the basis of the procedural justice and/or distributive justice and/or effectiveness and/or lawfulness displayed by police officers (using the sort of statistical analysis employed by Huq et al. 2017, Sunshine and Tyler 2003, and others).

Mirroring the reasoning of Tankebe, Sun and others, you start with a conceptual stance that procedural justice, distributive justice, effectiveness and lawfulness are possible sources of legitimacy. You obtain Sun et al.'s (2018) data, you fit a confirmatory factor model to test whether your challenge to the standard practice (i.e. Sun et al. 2018) is correct, and you obtain the results summarised in Fig. 4. You find that the measures of the four constructs (a) scale well, (b) can be represented as four latent variables, (c) are strongly and positively correlated and (d) that these correlations between the latent variables can be modelled according to a second-order factor.

Crucially, having a priori labelled the second-order factor 'possible sources of legitimacy', you argue that the finding constitutes empirical proof that they are possible sources of legitimacy and

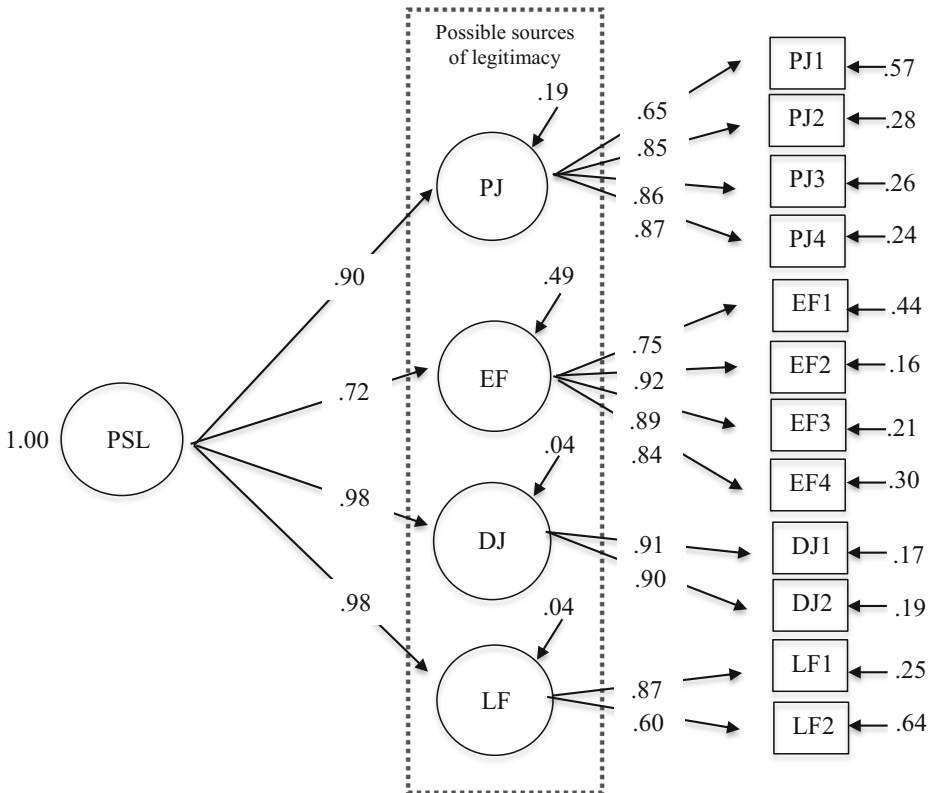


Fig. 4 An alternative measurement model of legitimacy. Source: Sun, I. Y., Li, L., Wu, Y., and Hu, R. (2018). Police legitimacy and citizen cooperation in China: testing an alternative model. *Asian Journal of Criminology*, 13(4), pp. 275–291. PSL = possible sources of legitimacy; PJ = procedural justice; EF = effectiveness; DJ = distributive justice; LF = lawfulness. Fit statistics: $\chi^2 = 206$, $df = 50$, $p < .005$; CFI = .982; TLI = .976; RMSEA = .058; and SRMR = .030. Because this model fitted the data reasonably well and because the measures had good scaling properties, you argue that the second-order factor must be labelled ‘possible sources of legitimacy’ and not ‘constituent components of legitimacy’

not constituent components of legitimacy. Note that this follows the line of reasoning of Tankebe (2013) and Sun et al. (2018). The fact that the model fits means that you were right to define procedural justice, distributive justice, effectiveness and lawfulness as possible sources of legitimacy (rather than constituent components of legitimacy). Having claimed that the finding overturns current thinking in this coastal Chinese city, you then move on to test one of two models linking the potential predictors of legitimacy to obligation to obey (Fig. 5).

Yet, imagine , for one moment, pausing to reflect. Tankebe (2013) and Sun et al. (2018) began by noting the status quo: namely, that procedural justice, distributive justice, effectiveness and lawfulness are possible sources of legitimacy (with regression modelling indicating the most important culturally contingent criteria). They then claimed that the findings from their CFA modelling showed that these four constructs were, in fact, constituent components of legitimacy. By arguing that the fitted CFA models in Figs. 2 and 3 were incompatible with the idea that the constructs were possible sources of legitimacy, they overturned the status quo. In the words of Sun et al. (2018: p. 14), procedural justice, distributive justice, effectiveness and lawfulness are ‘indicators, rather than antecedents of legitimacy’.

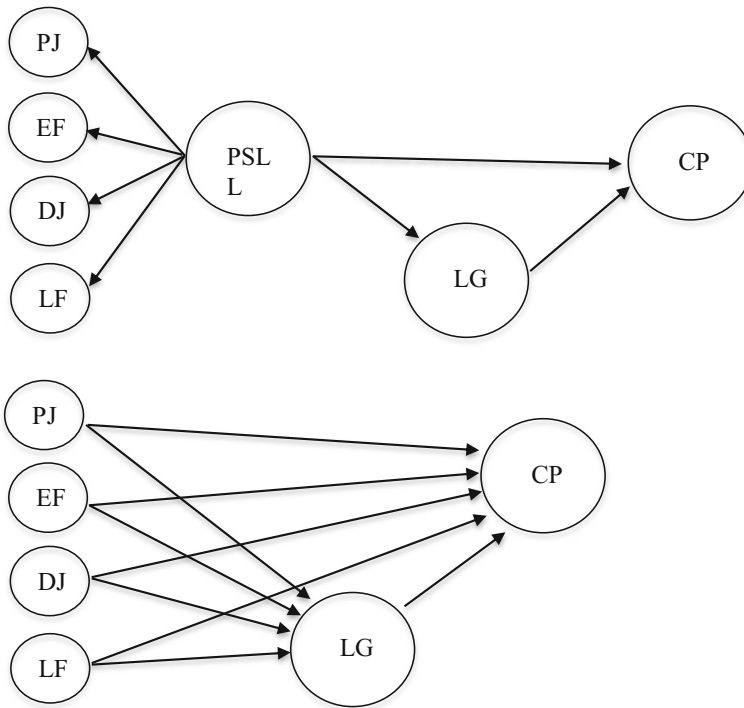


Fig. 5 Two theoretical models of possible sources of legitimacy, legitimacy and cooperation. PSL = possible sources of legitimacy; LG = legitimacy (measured using indicators of felt obligation to obey the police); CP = willingness to cooperate with the police

You are doing the opposite. You start with their status quo, i.e. that procedural justice, distributive justice, effectiveness and lawfulness are constituent components of legitimacy (Sun et al. 2018). Your CFA modelling shows good convergence validity, discriminant validity and internal consistency (to paraphrase Sun et al. 2018, p. 14). You find that a second-order factor (that you a priori labelled ‘possible sources of legitimacy’) model fitted the data. So, you argue that what have previously been treated as constituent components of legitimacy are in fact possible sources of legitimacy. You conclude, in other words, that you were correct *in the first place* in how you labelled the constructs, and that Sun et al. (2018) were incorrect *in the first place* when they labelled the constructs.

Yet, it is plainly apparent that Figs. 3 and 4 are identical, apart from the label given to the second-order factor. Tankebe (2013) viewed the fitted model (Fig. 2) to be incompatible with the notion that they are possible sources of legitimacy (as they are labelled in Huq et al. 2017; see Fig. 1). But if both conceptual bases are consistent with the data, how can CFA provide empirical evidence on which of the two competing conceptual stances is ‘correct’? In Sun et al.’s (2018) study, the model fits the data reasonably well, regardless of what label we assign to the second-order factor.

CFA is good at modelling correlations between variables according to some hypothesised latent structures; it is less adept at telling us what to label a collection of constructs. Whether one calls these constructs possible sources of legitimacy or constituent components of legitimacy depends on one’s conceptual stance. Indeed, to believe that CFA constitutes a test of the conceptual status of the constructs, one would have to reify latent variables in a quite unusual way. One would have to name

the second-order factor before fitting the model, then interpret the fact that a second-order factor model fits the data as empirical proof that one was right to name the second-order factor in the way that you did, prior to doing the analysis. Yet, the fitted model produces the same results whatever we call the second-order factor. So, why is it a test one way or the other?

Indeed, it is difficult to know what would have to be present in a fitted CFA model if the empirical test described above (testing whether procedural justice, distributive justice, effectiveness and lawfulness are constituent components of legitimacy) was to be deemed to fail. In Tankebe's (2013) analysis, it seems that if the fit of the model was reduced by adding effectiveness, then this proves that the police do not have to be seen to act effectively if they are to be seen as legitimate by citizens. But why would that follow? A less well fitting model would more likely indicate that there are cross-loadings and/or error covariances that should be added, not that one is wrong to label a collection of constructs in a particular way. In Sun et al.'s (2018) analysis, it could be that a single second-order factor would not fit the data, but why would finding that one needs multiple second-order factors to explain why procedural justice, distributive justice, effectiveness and lawfulness are correlated somehow 'prove' that procedural justice, distributive justice, effectiveness and lawfulness are not potential sources of legitimacy? There is no requirement that legitimacy is unidimensional.

Testing the Cultural (In)Sensitivity of This Approach to Measurement: a 30-Country Study

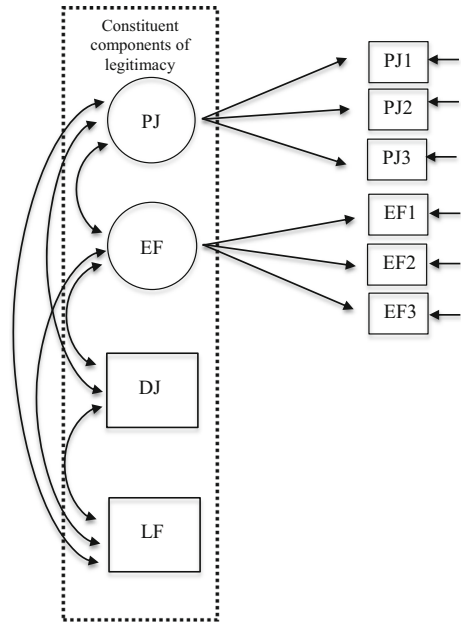
So far in this paper we have discussed what latent variable modelling can and cannot say about an issue of construct validity that is ultimately down to conceptual analysis and operational argumentation. In the second part, we turn to the idea that the measures of procedural justice, distributive justice, effectiveness and lawfulness are likely to scale well in most places one looks at; that if one (wrongly) interprets good scaling properties as evidence that these constructs are constituent components rather than possible sources of legitimacy, then one would (wrongly) conclude that legitimacy *is* procedural justice, distributive justice, effectiveness and lawfulness in most places one looks at; and that the approach lacks cultural sensitivity because it would end up asserting, without credible empirical evidence, that the normative preconditions of legitimacy are exactly the same in nearly all the places one looks at.

To reflect on the issue of cultural sensitivity, we apply the approach to 30 different countries. Our goal is to test whether the CFA model of Tankebe (2013) is likely to fit wherever one looks for two reasons: (a) we as a community of criminologists have good measures of these four constructs and (b) in most contexts the constructs are likely to be empirically distinct but positively correlated. Using the same approach to measuring and modelling legitimacy as Tankebe (2013), we fit the same CFA model in each country separately. We distinguish between procedural justice, distributive justice, effectiveness and lawfulness (Fig. 6), and following Tankebe (2013), we call these constituent components of legitimacy.

Data

The ESS is an academically driven face-to-face interview survey that runs every 2 years. Charting a range of attitudes, values, behaviours and beliefs between nations and over time, it is one of the highest quality—if not the highest quality—cross-national surveys in the world, especially in terms of sampling and measurement equivalence. It employs a rigorous questionnaire translation,

Fig. 6 Posited measurement model of potential sources of legitimacy. PJ = procedural justice; EF = effectiveness; DJ = distributive justice; LF = lawfulness



pretesting and development methodology (Jowell et al. 2007). Although not all countries achieve it, the aspiration is that countries should have probability samples of the adult (16+) population, with high response rates, interviewed face-to-face using CAPI (computer-assisted personal interviewing). The questionnaire comprises an invariant core of questions asked of all respondents in each round. Also included in some rounds are rotating modules that focus in detail on a particular issue. Academics are invited to bid for space on the questionnaire in each round.

In Round 5, a module on *trust in justice* containing 45 questions was included (European Social Survey 2011; Jackson et al. 2011; Hough et al. 2013a). Fieldwork for Round 5 of the ESS was done in 2010/2011 (European Social Survey, 2010, 2018). A total of 28 countries took part, some of which were *European* in only a loose sense: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Lithuania, Netherlands, Norway, Poland, Portugal, Russian Federation, Slovakia, Slovenia, Spain, Sweden, Switzerland, UK and Ukraine. The probability samples are representative of all persons aged 15 and older, and residents within the borders of the nation, regardless of nationality, citizenship, language or legal status. The smallest sample size was 1083 in Cyprus and the largest sample size was 3031 in Germany.

The US data come from an internet-based survey fielded to a random selection of individuals drawn from a GFK Knowledge Networks research panel of U.S. adults (Tyler and Jackson 2014; Tyler et al. 2015). Knowledge Networks uses random digit dialling and address-based sampling methods to construct and maintain the panel. A total of 2561 respondents were initially selected from the larger panel. The study was described, an offer of compensation extended and a reminder email was sent to all people on the list who had not responded after 3 days. The survey was fielded in August and September of 2012, either in English or in Spanish. A total of 1603 individuals completed the survey, representing a response rate of 62.5% from the existing internet panel.

The South Africa data come from the 2012 round of the South African Social Attitudes Survey (SASAS), a repeated cross-sectional survey conducted annually by the Human Sciences Research Council. The survey round consisted of a nationally representative probability sample of 3183 South African adults aged 16 years and over living in private households. Each SASAS round of interviewing consists of a sub-sample of 500 Population Census enumeration areas (EAs), stratified by province, geographical sub-type and majority population group. The SASAS aims to provide a long-term account of change in public values and the social fabric of modern South Africa. Given the importance of issues of crime and policing in South African society, permission was secured to field police-related questions from the trust in justice module included in the fifth round of the European Social Survey in 2010/2011 (see Bradford et al. 2014a, 2014b).

Measures

Procedural justice was measured by asking respondents how often (from 1 ‘not at all often’ to 4 ‘very often’) they think that officers in their country:

- Treat people with respect;
- Make fair and impartial decisions; and,
- Explain their decisions and actions when asked.

Police effectiveness was measured by asking respondents: ‘Based on what you have heard or your own experience, how successful do you think the police in this country are at...’ (on a scale from 0 to 10):

- Preventing crimes where violence is used or threatened;
- Catching people who commit house burglaries; and,
- How slowly or quickly police would arrive at the scene if a violent crime were to occur near your house.

Distributive justice was measured using two questions. The introduction was: ‘Now some questions about whether or not the police in [country] treat victims of crime equally. Please answer based on what you have heard or your own experience.’ The first question was: ‘When victims report crimes, do you think the police treat rich people worse, poor people worse, or are rich and poor treated equally?’ The second question was: ‘when victims report crimes, do you think the police treat some people worse because of their race or ethnic group or is everyone treated equally?’ These two indicators were combined to form a single variable, with 0 equalling ‘neither poor nor minority group members are treated worse’, 1 equalling ‘either poor or minority group members are treated worse’ and 2 equalling ‘both poor and minority group members are treated worse’. Finally, lawfulness was measured by asking people (a) to agree or disagree with the statement ‘Decisions and actions of police are unduly influenced by political pressure’ on a 5-point Likert scale and (b) ‘How often do police [in your country] take bribes’ on a scale from 0 ‘never’ to 10 ‘always’. These two variables were rescaled, such that high scores equal the belief that the police act lawfully. A single index was created by taking the mean of the two variables (having divided the bribery variable by 2 to put it on a comparable scale to the political pressure variable).

Note that a strength of the data is that the same measures were fielded in nationally representative sample surveys of 30 countries. A weakness is that the measures of distributive justice and lawfulness were limited to two each. The measures of distributive justice were nominal

with three categories each, and the measures of lawfulness were one five-category variable and one ten-category variable. As such, it makes sense to produce a single derived variable for distributive justice and a single derived variable for lawfulness. In the context of confirmatory factor analysis, this means we cannot test a model exactly like Tankebe's (2013). But it is close enough because the CFA is still assessing, among other things, whether these are four empirically distinct constructs (Fig. 6). By close enough, we mean that whether one (a) has three indicators of distributive justice, three indicators of lawfulness and models each as a latent variable or (b) has two indicators of distributive justice, two indicators of lawfulness, and combines each to create a manifest variable will make only a little difference to the fit statistics (improving them a little bit because there are fewer constraints in the model to test) and this should make only a marginal difference in the context of what we are trying to test here. More important is the extent to which the four constructs are empirically distinct and positively correlated.

Note, also, that we do not fit a second-order factor. The extra testable assumption at the heart of a second-order factor is unnecessary; it involves testing whether the correlations between the four latent constructs can be modelled according to a single underlying second-order latent construct, but as detailed above, if one finds that the second-order factor model fits, why would this increase one's confidence that the four constructs are constituent components of legitimacy rather than possible sources of legitimacy? Because the researcher is free to call the second-order factor whatever she likes? Equally, Sun et al. (2018) sought to replicate Tankebe (2013) and he presented legitimacy as multidimensional and did not fit a second-order factor, so ours is a more direct replication in that sense.

Results

Table 1 provides the exact and approximate fit statistics CFA models estimated for each country in the dataset (Fig. 6) that differentiates between procedural justice (a latent construct with three indicators), effectiveness (another latent construct with three indicators), distributive justice (a manifest indicator calculated using responses to two indicators) and lawfulness (another manifest indicator calculated using responses to two indicators). We discount the exact fit statistics because the chi-square test is extremely sensitive to sample size (very small deviations in the match between the hypothesised and saturated models can be highlighted as statistically significant). The approximate fit statistics are adequate in all countries (cf. Hu and Bentler 1998). The root mean square error of approximation (RMSEA) statistics are all close to the cut-off point of .06; RMSEA is a parsimony-adjusted index, with values closer to 0 indicating a good fit. The comparative fit index (CFI) statistics are all above .95; CFI compares the fit of the model to the fit of a null model, with values closer to 1 indicating a good fit. The Tucker Lewis Index (TLI) statistics are all close to the cut-off point of .95; TLI is an incremental measure of fit that has a penalty for adding parameters, with values closer to 1 indicating a good fit.

Table 2 provides details of the scaling properties of procedural justice and effectiveness in each country. Each cell gives the standardised factor loading (left) and R^2 (right) for each particular indicator (PJ1, PJ2, PJ3, EFF1, EFF2 and EFF3) in each of the 30 different countries. For procedural justice, the standardised factor loadings range from .59 to .94, and the R^2 's range from .35 to .88. For effectiveness, the standardised factor loadings range from .44 to .90, and the R^2 's range from .20 to .80. Because the factor loadings and R^2 's are all relatively high, this indicates good scaling properties.

Table 3 provides the correlations between the four constructs. Looking across the 30 countries, the pair of constructs with the strongest correlation is procedural justice and

Table 1 Fit statistics for the fitted confirmatory factor analysis model in each of the 30 countries

Country	Chi-square	df	<i>p</i> value	RMSEA 95% CI	CFI	TLI
Austria	149	16	<0.001	.061 [.052, .070]	.987	.978
Belgium	146	16	<0.001	.069 [.059, .080]	.956	.923
Bulgaria	110	16	<0.001	.050 [.041, .059]	.992	.986
Croatia	154	16	<0.001	.073 [.063, .083]	.973	.953
Cyprus	69	16	<0.001	.055 [.042, .069]	.991	.985
Czech Republic	157	16	<0.001	.061 [.063, .070]	.982	.968
Denmark	103	16	<0.001	.059 [.048, .070]	.968	.945
Estonia	91	16	<0.001	.051 [.042, .062]	.979	.963
Finland	78	16	<0.001	.046 [.036, .056]	.980	.965
France	120	16	<0.001	.061 [.051, .072]	.974	.954
Germany	143	16	<0.001	.051 [.044, .059]	.974	.955
Greece	99	16	<0.001	.044 [.036, .052]	.995	.991
Hungary	51	16	<0.001	.038 [.026, .050]	.989	.981
Ireland	142	16	<0.001	.055 [.047, .064]	.986	.975
Israel	184	16	<0.001	.068 [.060, .077]	.980	.965
Lithuania	74	16	<0.001	.047 [.036, .058]	.986	.976
Netherlands	129	16	<0.001	.062 [.052, .072]	.962	.933
Norway	131	16	<0.001	.068 [.058, .079]	.957	.924
Poland	83	16	<0.001	.049 [.039, .060]	.985	.974
Portugal	186	16	<0.001	.070 [.061, .080]	.973	.953
Russian Federation	174	16	<0.001	.062 [.054, .071]	.984	.971
Slovakia	139	16	<0.001	.065 [.055, .075]	.982	.969
Slovenia	93	16	<0.001	.059 [.047, .071]	.982	.968
South Africa	182	16	<0.001	.065 [.056, .073]	.986	.976
Spain	50	16	<0.001	.033 [.023, .044]	.994	.990
Sweden	101	16	<0.001	.060 [.049, .071]	.963	.935
Switzerland	70	16	<0.001	.047 [.036, .059]	.978	.962
Ukraine	95	16	<0.001	.051 [.041, .061]	.990	.982
UK	99	16	<0.001	.046 [.038, .055]	.988	.979
USA	96	16	<0.001	.056 [.046, .067]	.989	.981

Data: Round 5 European Social Survey, Yale Trust in Law Survey and SASAS 2010

effectiveness (ranges from .45 to .77) and the pair of constructs with the weakest correlation is distributive justice and lawfulness (ranges from .06 to .42).

If we applied the reasoning of Tankebe (2013) and Sun et al. (2018) with respect to their CFA modelling, we would interpret the fact that the four-factor model fitted the data in each of the 30 countries as evidence that what have been previously treated as possible sources of legitimacy are, in fact, constituent components of legitimacy. This would imply that legitimacy rests on the same normative bases (in the sense that people judge the legitimacy of the police using the same defining criteria of appropriate police conduct) in all 30 countries. In each of these diverse social, political and legal contexts, the police need to be seen to act in ways that are procedurally just, distributively just, effective and lawful if they are to be deemed empirically legitimate.

Strikingly, this would contradict a good deal of existing research that (a) shows that procedural justice is the most important criterion in many countries across the world, and (b) highlights some interesting country-level differences in the extent to which each of the four criteria explains variation in legitimacy (see Jackson 2018, for a review of the international literature). Indeed, the claim here is not that procedural justice, distributive justice, effectiveness and lawfulness are all strong predictors of legitimacy in each country; the point is that they are so fundamental to the perceived right to power that they collectively *constitute* it. We return to this point in the discussion below.

Table 2 Standardised factor loadings and R²s for the PJ and EFF indicators

Country	PJ1	PJ2	PJ3	EFF1	EFF2	EFF3
Austria	.88 / .77	.93 / .87	.72 / .52	.80 / .65	.78 / .61	.78 / .60
Belgium	.78 / .61	.73 / .53	.65 / .42	.75 / .56	.70 / .49	.52 / .28
Bulgaria	.90 / .81	.93 / .87	.80 / .63	.85 / .72	.79 / .63	.71 / .51
Croatia	.84 / .71	.83 / .69	.78 / .60	.86 / .75	.83 / .69	.70 / .48
Cyprus	.88 / .77	.93 / .86	.80 / .64	.83 / .69	.79 / .63	.75 / .56
Czech Republic	.88 / .78	.88 / .77	.71 / .50	.78 / .61	.68 / .46	.70 / .50
Denmark	.77 / .59	.80 / .64	.70 / .48	.77 / .59	.53 / .28	.56 / .31
Estonia	.81 / .65	.81 / .65	.65 / .43	.70 / .49	.67 / .45	.61 / .37
Finland	.74 / .55	.80 / .64	.65 / .43	.70 / .49	.70 / .49	.54 / .29
France	.77 / .60	.84 / .70	.63 / .40	.77 / .59	.74 / .54	.63 / .39
Germany	.74 / .55	.82 / .67	.59 / .35	.67 / .45	.66 / .43	.55 / .30
Greece	.89 / .79	.94 / .88	.77 / .60	.86 / .74	.90 / .80	.81 / .66
Hungary	.82 / .68	.87 / .75	.62 / .38	.80 / .64	.75 / .56	.57 / .33
Ireland	.83 / .69	.91 / .82	.74 / .54	.80 / .63	.69 / .47	.66 / .43
Israel	.86 / .74	.86 / .73	.80 / .63	.78 / .60	.68 / .47	.61 / .37
Lithuania	.75 / .56	.83 / .68	.78 / .60	.83 / .69	.79 / .63	.64 / .41
Netherlands	.78 / .61	.79 / .62	.62 / .39	.71 / .50	.64 / .41	.56 / .32
Norway	.85 / .72	.68 / .47	.62 / .38	.73 / .53	.69 / .48	.62 / .39
Poland	.81 / .66	.84 / .71	.76 / .58	.81 / .65	.78 / .61	.65 / .42
Portugal	.77 / .60	.87 / .76	.79 / .62	.77 / .59	.78 / .61	.73 / .53
Russian Federation	.84 / .70	.89 / .80	.77 / .59	.81 / .66	.80 / .64	.70 / .48
Slovakia	.83 / .69	.91 / .83	.76 / .58	.83 / .69	.79 / .62	.65 / .43
Slovenia	.83 / .69	.92 / .85	.67 / .45	.80 / .63	.78 / .60	.62 / .38
South Africa	.87 / .76	.91 / .83	.79 / .62	.83 / .69	.78 / .61	.80 / .64
Spain	.84 / .71	.89 / .79	.70 / .49	.82 / .67	.76 / .57	.66 / .44
Sweden	.70 / .48	.73 / .53	.70 / .49	.76 / .58	.70 / .50	.56 / .31
Switzerland	.76 / .58	.76 / .58	.68 / .46	.74 / .55	.78 / .61	.44 / .20
Ukraine	.83 / .69	.92 / .84	.84 / .70	.84 / .70	.84 / .70	.75 / .56
UK	.86 / .73	.87 / .76	.70 / .49	.75 / .56	.71 / .50	.64 / .41
USA	.92 / .84	.79 / .63	.84 / .70	.85 / .71	.73 / .53	.72 / .52

Data: Round 5 European Social Survey, Yale Trust in Law Survey and SASAS 2010

Discussion

We began this paper by describing how the empirical concept of legitimacy (Hinsch 2008) specifies the right to power as a property of public opinion, e.g. people in Japan may view their police to be more legitimate than people in the Russian Federation view their police, and if this is the case then empirical legitimacy would be higher in Japan than in the Russian Federation. It also treats as an empirical question the culturally contingent criteria that citizens of a given country use to judge the legitimacy of the institution, e.g. people in Japan may judge the legitimacy of the police according to a different set of defining criteria of appropriate police conduct compared to people in the Russian Federation (i.e. the norms that specify rightful police behaviour might be different in the two countries). This is in contrast to the normative concept of legitimacy, which typically involves an outsider observer determining the same substantive requirements for legitimacy in different contexts. It could be, for example, that outside experts judge police legitimacy in both Japan and Russian Federation according to the same system-level criteria, e.g. independence, accountability and other indicators of the rule of law. To get a measure of normative legitimacy, these experts could collect national-level indicators of institutional arrangements and practices defined according to these criteria.

Table 3 Correlations between constructs from the fitted measurement model

Country	PJ with EFF	PJ with DJ	PJ with LF	EFF with DJ	EFF with LF	DJ with LF
Austria	.63	.54	.43	.47	.42	.42
Belgium	.51	.32	.36	.23	.28	.19
Bulgaria	.76	.49	.55	.47	.52	.42
Croatia	.71	.54	.48	.49	.40	.32
Cyprus	.70	.44	.52	.41	.46	.34
Czech Republic	.70	.46	.53	.39	.39	.31
Denmark	.58	.38	.27	.35	.25	.22
Estonia	.67	.39	.47	.34	.41	.35
Finland	.52	.32	.38	.27	.21	.25
France	.64	.45	.40	.30	.38	.32
Germany	.55	.35	.32	.36	.29	.28
Greece	.77	.40	.56	.37	.49	.33
Hungary	.65	.47	.51	.40	.38	.37
Ireland	.68	.41	.43	.42	.39	.33
Israel	.71	.40	.32	.38	.27	.19
Lithuania	.69	.42	.45	.41	.36	.29
Netherlands	.45	.38	.29	.26	.16	.19
Norway	.53	.31	.31	.29	.16	.17
Poland	.63	.46	.47	.41	.41	.37
Portugal	.51	.41	.37	.44	.47	.34
Russian Federation	.75	.42	.41	.40	.36	.25
Slovakia	.64	.39	.45	.40	.42	.29
Slovenia	.64	.43	.42	.43	.36	.36
South Africa	.54	.43	.33	.41	.27	.24
Spain	.68	.45	.45	.40	.41	.34
Sweden	.55	.33	.29	.34	.09	.18
Switzerland	.52	.36	.34	.30	.25	.29
Ukraine	.69	.30	.31	.32	.25	.06
UK	.62	.38	.35	.37	.33	.28
USA	.67	.48	.61	.34	.49	.40

Data: Round 5 European Social Survey, Yale Trust in Law Survey and SASAS 2010

We then described the methodology used by Sun et al. (2018) and Tankebe (2013) to test a new approach to defining, measuring and modelling empirical legitimacy. In Sun et al.'s (2018) study, the scales of procedural justice, distributive justice, effectiveness and lawfulness had good measurement properties; they loaded on four strongly correlated latent variables; and these four latent constructs were themselves regressed onto a single second-order factor (that the researchers labelled legitimacy). The researchers subsequently argued that, on this evidence base, they are such *strong* criteria that they can be treated as constituent components of legitimacy. What have previously been treated as potential forms of legitimation—types of normatively appropriate police conduct—are now reclassified as legitimacy beliefs themselves.

We then took the reader through a hypothetical reanalysis of Sun et al.'s (2018) data. Taking a different starting point but following the same logical sequence, we reached the conclusion that what has previously been viewed as four constituent components of legitimacy in this coastal Chinese city are, in fact, possible sources of legitimacy. The scales of procedural justice, distributive justice, effectiveness and lawfulness had good measurement properties; they loaded on four strongly correlated latent variables; and these four latent constructs were themselves regressed onto a single second-order factor one a priori labelled 'possible sources of legitimacy'. This thought experiment illustrated the basic methodological point that the findings of the CFA modelling do not tell whether one is right *in the first place* to define

procedural justice, distributive justice, effectiveness and lawfulness as (a) potential criteria that people use to judge legitimacy or (b) constituent components of legitimacy. Put bluntly, CFA is not a good adjudication tool because the same fitted model is consistent with both conceptual stances; we get exactly the same findings *whatever we call the collection of constructs*. As a result, we recommend that researchers do not use CFA in the same way as Tankebe (2013) and Sun et al. (2018).

We then investigated whether the approach to defining, measuring and modelling police legitimacy of Sun et al. (2018) and Tankebe (2013) ironically does the opposite of embracing cultural diversity. By linking Round 5 of the European Social Survey (ESS) to two matching representative sample surveys of the USA and South Africa, we produced a 30-country dataset spanning a range of diverse contexts. Testing whether Tankebe's (2013) four-factor model fitted in each of the social, political and legal contexts, we found that the model (specified in Fig. 6) fitted the data reasonably well in each and every case. In each and every country, measures of procedural justice, distributive justice, effectiveness and lawfulness scaled reasonably well; they reflected or formed four empirically distinct constructs; and they were strongly and positively correlated with each other. Following the reasoning of Sun and colleagues, this would imply that legitimacy is—in each and every one of those countries—the same thing, i.e. constituted by public judgements of police procedural justice, distributive justice, effectiveness and lawfulness. In Sun et al.'s (2018) view, the legitimating norms that people expect police to abide by, before the institution is to be viewed as legitimate, would be the same in all countries and relate to these four types of police behaviour.¹

¹ We should say that an anonymous referee asked whether it really matters whether criminologists define procedural justice, distributive justice, effectiveness and lawfulness judgements as potential sources of legitimacy or constituent components of legitimacy. To quote the referee: 'The difference between possible sources of legitimacy and constituent components of legitimacy might be real as the difference between the interval level measurement and the ratio level of measurement, but criminology as a soft science does not care much about that difference. We constantly treat ordinal measures as if they were interval/ratio.' We are happy to clarify our position. When criminologists study the legitimacy of the police, they tend to try to answer two questions. First, what legitimates the police in the eyes of citizens, i.e. what types of police conduct legitimate or delegitimize the institution in the eyes of the policed? Second, does legitimacy motivate public cooperation and compliance, i.e. does legitimacy motivate people to comply with norms and rules? The first line of enquiry is central to the current paper. Observational studies model the correlations between people's contact with the police, their perceptions of how officers generally act (do officers generally act in procedurally just, distributively just, effective and lawful ways?) and legitimacy (overarching judgements of the institution's right to power and authority to govern). Experimental studies manipulate police behaviour and assess whether legitimacy is 'moved around' (e.g. Posch, 2019; Trinkner et al. 2019). Now, if we as a community of criminologists decide that it does not matter whether procedural justice, distributive justice, effectiveness and lawfulness are potential sources of legitimacy or constituent components of legitimacy, then that would shut down survey and experimental work into the criteria that people use to judge the legitimacy of the police (at least in terms of procedural justice, distributive justice, effectiveness and lawfulness). While the anonymous referee suggests that this may be fine, we respectfully disagree. We think it is important to embrace cultural diversity in the criteria that people use to judge rightfulness. Moreover, to make evidence-based policy recommendations on how the police should act to encourage legitimacy, one needs empirical research. It may be that in the USA, UK and Australia, for instance, people base their judgements of the legitimacy of the police as an institution to a significant extent on whether officers act in procedurally fair ways, but in countries like South Africa and Pakistan the effectiveness of the police in fighting crime is as least as important as procedural justice. If, as the anonymous referee seems to think, it is really just semantics whether we define procedural justice, distributive justice, effectiveness and lawfulness as possible sources of legitimacy or constituent components of legitimacy (as important as the 'difference between the interval level of measurement and the ratio level of measurement'), then it would be impossible to study empirically what legitimates or delegitimizes the police, and it would be difficult to draw empirically informed policy recommendations. To do all this, one needs, by definition, to tease apart rather than conflate potential predictors of legitimacy and constituent components of legitimacy.

We believe this would be a misfounded view. The fact that the model fitted the data well in each country says little about whether we are measuring, accurately or not, legitimacy. The claim that we are doing so is a purely conceptual matter based on an a priori assumption about what constitutes legitimacy in a particular context and how it can reasonably be operationalised. Moreover, the approach would lead to what seems to us the rather odd conclusion that the criteria citizens use to judge police legitimacy are exactly the same across all 30 countries (indeed so strong that they collectively constitute the construct itself) based erroneously on the fact that the measures had good scaling properties in each of the 30 countries.

If measures of procedural justice, distributive justice, effectiveness and lawfulness will scale well wherever one looks, then one would end up saying that legitimacy is defined as perceived procedural justice, distributive justice, effectiveness and lawfulness wherever one looks. In practice, therefore, far from being sensitive to cultural variation in the composition of legitimacy, the model proposed by Tankebe, Sun and colleagues flattens out the possibility of variation. It imposes an Anglo-Saxon perspective under the smokescreen of empirical discovery. We do not want this mistake to catch on—and given publication bias, it is more likely to catch on *if the measurement model fits nearly everywhere we look*.

There are three obvious problems with this. First, one could take a normative view that police should act in procedurally just, distributively just, effective and lawful ways, and that being seen to do so is so fundamental to their principled justification that their right to power rests directly on public perceptions across these criteria. Irrespective of the criteria that citizens *actually use* to judge legitimacy, this approach expounds the criteria that must be in place if the police are to be seen as legitimate from a normative sense. It employs one aspect of the normative concept of legitimacy—in that it is the outside expert (not the people in a given society) who decides what needs to be present for the police to be deemed legitimate—and fuses it to one aspect of the empirical concept of legitimacy, where legitimacy is defined as a judgement that citizens make regarding the right to power and authority to govern. This may be a reasonable position, albeit it is not to our own particular taste—we prefer to empirically discover the culturally contingent criteria of legitimation that people in a particular political community actually use. The problem lies, instead, with researchers claiming spurious empirical evidence for what is ultimately a normative position. This is not a question of whether our research should be theory-driven or data-driven (as one anonymous referee put it), it is a methodological question about what sort of evidence CFA does and does not provide.

Second, and relatedly, when using the approach to defining, measuring and modelling police legitimacy of Sun et al. (2018), there is little or no possibility of assessing which, if any, is the most important component of legitimacy. What type or aspect of behaviour is most important in generating a sense that that police activity is normatively justifiable? Do people value procedural justice most? Or are they more concerned with effectiveness? These seem to us important questions, both theoretically and from a policy perspective. Answers to these questions may vary across different cultural contexts, yet the approach to measuring legitimacy taken in Sun et al. (2018) makes it difficult if not impossible to answer them. If one defined the four constructs as legitimacy and erroneously interpreted CFA results as proving one is correct, then by definition it cannot be an empirical question whether one (or more) of these judgements is the most important to some overarching judgement of the right to power. Comparing the factor loadings of the four first-order latent constructs would not provide much that is informative because there is no requirement that legitimacy is unidimensional. One has *before the fact* decided that they are so important to legitimacy that they collectively constitute the very judgement itself.

The third problem with the approach taken by Sun and colleagues is that it leaves no room for the idea that other judgements might come into play when people are thinking about the normative appropriateness of police activity. These might be many and varied, and some might be morally troubling from a normative perspective. Some white US citizens, for example, might believe police are behaving appropriately when they target black US citizens, not because they think this makes policing more effective or fair—although they might also believe this—but because they have been socialised or otherwise come to believe this is just the way *police should behave*. Given the history of the USA vis-à-vis some other liberal democracies (Alexander 2012), there is no necessary reason to assume this would be the case elsewhere, although of course it might be. This is, to our minds, an empirical question worthy of investigation.

But the measurement model specified in Sun et al. (2018) renders such investigation analytically and conceptually difficult because it defines legitimacy only as procedural justice, distributive justice, effectiveness and lawfulness. One would have to assume, for example, that racially targeted policing in a given context is only a source of legitimacy to the extent that it influences beliefs about fairness, effectiveness and lawfulness. If it does not, then it has no effect on legitimacy. Indeed, *nothing* could have an effect on legitimacy if it did not have an effect on at least one of the four constructs. And, if one a priori defines procedural justice, distributive justice, effectiveness and lawfulness as legitimacy, then legitimacy cannot be constituted by anything else (otherwise one would have defined it as that too).

We counsel against Sun et al.'s (2018) study starting a trend in criminology, where researchers in a novel context use the same approach to defining and measuring police legitimacy as Tankebe (2013), use CFA in the same way as Tankebe (2013) and conclude from their findings that there is empirical evidence that the story that emerges from the novel context is different from that in the USA, UK and Australia. Studies like Sun et al. (2018) may pop up everywhere, find 'supportive evidence' because the items will scale well nearly everywhere that one looks, and make the same mistake of concluding that legitimacy is procedural justice, distributive justice, effectiveness and lawfulness on the basis of the fitted CFA model. Dressed up as an empirical finding, the approach does the opposite of 'embracing cultural diversity', to paraphrase Sun et al. (2018: p. 5). It assumes beforehand that these judgements concern procedural justice, distributive justice, effectiveness and lawfulness. So, we have what is a curious mix of the normative concept and the empirical concept of legitimacy. Public opinion matters with regard to overall levels of legitimacy in a given society, but it is the researcher who is imposing the substantive requirements for empirical legitimacy.

The answer to all these problems is obvious and is already employed in much of the literature. If legitimacy is conceptualised and measured as something distinct from assessments of officer fairness, effectiveness and lawfulness (i.e. a more overarching judgement of the institution's right to power and authority to govern), then it is possible to assess which if any of these is most important as a predictor of legitimacy. Similarly, if legitimacy is distinct and different from the four factors, then the influence of other variables is conceptually and analytically far easier to assess, since other judgements of the normative appropriateness of police activity are allowed to have effects distinct from any correlation with perceptions of fairness, effectiveness and lawfulness. For instance, respecting the limits of one's rightful authority may be important to legitimacy above and beyond perceptions of the procedural justice of the police (Huq et al. 2017; Trinkner et al. 2018).

To close, we do think there is space for alternative approaches to measuring legitimacy. Legitimacy is an abstract and unobservable psychological construct, and there are numerous ways to operationalise the perceived right to power, aside from the standard ways of

institutional trust and/or normative alignment and/or obligation to obey (Tyler and Jackson 2013). But researchers need to be aware of the limits of different approaches when trying to decide which approach to measurement is preferable. In the current case, the methodology of Sun and colleagues means ironically imposing *top-down* rather than discovering *bottom-up* the normative content of legitimacy. There may be normative and/or theoretical and/or practical reasons to decide this. But it is, in our view, crucial to be clear that this is what one is doing, especially when the analytical strategy offers little empirical support for the central conceptual claim.

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

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

- ES/L011611/1. *Legal Norms and Crime Control: A Comparative, Cross-national Analysis*

Funding for the three projects were provided by:

- European Commission and various national funding bodies on the European Social Survey
- Grants obtained by Professor Tom Tyler from Yale Law School and New York University
- South Africa Human Sciences Research Council

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent All participants provided informed consent.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Akinlabi, O. M. & Murphy, K. (2018). Dull compulsion or perceived legitimacy? Assessing why people comply with the law in Nigeria. *Police Practice and Research*, online advance access.
- Bradford, B. (2014). Policing and social identity: procedural justice, inclusion and cooperation between police and public. *Policing and Society*, 24(1), 22–43.
- Bradford, B., & Jackson, J. (2018). Police legitimacy among immigrants in Europe: institutional frames and group position. *European Journal of Criminology*, 15(5), 567–588.
- Bradford, B., Huq, A., Jackson, J., & Roberts, B. (2014a). What price fairness when security is at stake? Police legitimacy in South Africa. *Regulation & governance*, 8(2), 246–268.
- Bradford, B., Murphy, K., & Jackson, J. (2014b). Officers as mirrors: policing, procedural justice and the (re)production of social identity. *British Journal of Criminology*, 54(4), 527–550.
- Caldeira, G. A., & Gibson, J. L. (1995). The legitimacy of the Court of Justice in the European Union: models of institutional support. *American Political Science Review*, 89(2), 356–376.

- Dirikx, A., & Van den Bulck, J. (2013). Media use and the process-based model for police cooperation: an integrative approach towards explaining adolescents' intentions to cooperate with the police. *British Journal of Criminology*, 54(2), 344–365.
- European Social Survey (2011). Trust in justice: Topline findings from the European Social Survey, ESS Topline Results Series Issue 1. By Jackson, J., Hough, M., Bradford, B., Pooler, T. M., Hohl, K. & Kuha, J.
- European Social Survey Round 5 (2010). Data file edition 3.4. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- European Social Survey Round 5 (2018): ESS-5 2010 Documentation report. Edition 4.2. Bergen, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC.
- Gerber, M. M., Gonzalez, R., Carvacho, H., Jimenez-Moya, G., Moray, C., & Jackson, J. (2018). On the justification of intergroup violence: the roles of procedural justice, police legitimacy and group identity in attitudes towards violence among indigenous people. *Psychology of Violence*, 8(3), 379–389.
- Gibson, J. L., Caldeira, G. A., & Spence, L. K. (2003). Measuring attitudes toward the United States Supreme Court. *American Journal of Political Science*, 47(2), 354–367.
- Hamm, J. A., Trinkner, R., & Carr, J. D. (2017). Fair process, trust, and cooperation: moving toward an integrated framework of police legitimacy. *Criminal Justice and Behavior*, 44(9), 1183–1212.
- Hinsch, W. (2008). Legitimacy and justice. In J. Kuhnelt (Ed.), *Political legitimization without morality?* London: Springer.
- Hinsch, W. (2010). Justice, legitimacy, and constitutional rights. *Critical Review of International Social and Political Philosophy*, 13, 39–54.
- Hough, M., Jackson, J., & Bradford, B. (2013a). The governance of criminal justice, legitimacy and trust. In S. Body-Gendrot, R. Lévy, M. Hough, S. Snacken, & K. Kerezi (Eds.), *Routledge handbook of European criminology* (pp. 243–265). Oxon: Routledge.
- Hough, M., Jackson, J., & Bradford, B. (2013b). Legitimacy, trust and compliance: an empirical test of procedural justice theory using the European Social Survey. In J. Tankebe & A. Liebling (Eds.), *Legitimacy and criminal justice: an international exploration* (pp. 326–352). Oxford: Oxford University Press.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453.
- Huq, A. Z., Tyler, T. R., & Schulhofer, S. J. (2011). Mechanisms for eliciting cooperation in counterterrorism policing: a study of British Muslims. *Journal of Empirical Legal Studies*, 8, 728–761.
- Huq, A. Z., Jackson, J., & Trinkner, R. (2017). Legitimizing practices: revisiting the predicates of police legitimacy. *British Journal of Criminology*, 57(5), 1101–1122.
- Jackson, J., & Gau, J. (2015). Carving up concepts? Differentiating between trust and legitimacy in public attitudes towards legal authority. In E. Shockley, T. M. S. Neal, L. PytlikZillig, & B. Bornstein (Eds.), *Interdisciplinary perspectives on trust: towards theoretical and methodological integration* (pp. 49–69). New York: Springer.
- Jackson, J. (2018). Norms, normativity and the legitimacy of legal authorities: international perspectives. *Annual Review of Law and Social Science*, 14, 145–165.
- Jackson, J., Asif, M., Bradford, B., & Zakar, M. Z. (2014). Corruption and police legitimacy in Lahore, Pakistan. *British Journal of Criminology*, 54, 1067–1088.
- Jackson, J., Bradford, B., Hough, M., Kuha, J., Stares, S. R., Widdop, S., Fitzgerald, R., Yordanova, M., & Galev, T. (2011). Developing European indicators of trust in justice. *European Journal of Criminology*, 8(4), 267–285.
- Jackson, J., Bradford, B., Stanko, E. A., & Hohl, K. (2013). *Just authority? Trust in the police in England and Wales*. Oxon: Routledge.
- Jowell, R., Roberts, C., Fitzgerald, R., & Eva, G. (Eds.). (2007). *Measuring attitudes cross-nationally: lessons from the European Social Survey*. London: Sage.
- Justice, B., & Meares, T. (2014). How the criminal justice system educates citizens. *Annals of the American Academy of Political and Social Sciences*, 651(1), 159–177.
- Kim, Y. S., Ra, K. H. & McLean, K. (2018). The generalizability of police legitimacy: procedural justice, legitimacy, and speeding intention of South Korean drivers. *Asian Journal of Criminology*, advance online access.
- Liebling, A. (2004). *Prisons and their moral performance*. Oxford: Oxford University Press.
- Meares, T. (2017). Policing and procedural justice: shaping citizens' identities to increase democratic participation. *Northwestern University Law Review*, 111(6), 1525–1536.
- Mentovich, A., Ben-Porat, G., Levy, N., Goff, P. A. & Tyler, T. (2018). Policing alienated minorities in divided cities. *Regulation & Governance*, advance online access.
- Murphy, K., & Cherney, A. (2012). Understanding cooperation with police in a diverse society. *British Journal of Criminology*, 52(1), 181–201.

- Murphy, T., Bradford, B., & Jackson, J. (2016). Motivating compliance behavior among offenders: procedural justice or deterrence? *Criminal Justice and Behavior*, 43(1), 102–118.
- Murphy, K., Chemey, A., & Teston, M. (2018). Promoting Muslims' willingness to report terror threats to police: testing competing theories of procedural justice. *Justice Quarterly*, advance online access, 1–26.
- Oliveira, T. R., Oliveira, A. R. & Adorno, S. (2019). Legitimidade da polícia: Um modelo de mensuração. *Revista Brasileira de Ciências Sociais*, 34(100).
- Pardo, I. (2000). *Morals of legitimacy: between agency and system*. New York: Berghahn Books.
- Pösch, K. (2019). Testing complex social theories with causal mediation analysis and G-computation: toward a better way to do causal structural equation modeling. *Sociological Methods & Research*, advance access, 0049124119826159.
- Pösch, K., Jackson, J., Bradford, B. and MacQueen, S. (2019). Truly free consent? Clarifying the nature of police legitimacy using causal mediation analysis (May 14, 2019). Available at SSRN: <https://doi.org/10.2139/ssrn.2620274><https://ssrn.com/abstract=2620274>. Accessed 14 May 2019
- Reisig, M. D., Bratton, J. & Gertz, M. G. (2007). The construct validity and refinement of process-based policing measures. *Criminal Justice Behavior*, 34, 1005–1027.
- Sun, I. Y., Li, L., Wu, Y., & Hu, R. (2018). Police legitimacy and citizen cooperation in China: testing an alternative model. *Asian Journal of Criminology*, 13(4), 275–291.
- Sun, I. Y., Wu, Y., Hu, R., & Farmer, A. K. (2017). Procedural justice, legitimacy, and public cooperation with police: does Western wisdom hold in China? *Journal of Research in Crime and Delinquency*, 54(4), 454–478.
- Sunshine, J., & Tyler, T. R. (2003). The role of procedural justice and legitimacy in public support for policing. *Law & Society Review*, 37, 513–548.
- Tankebe, J. (2013). Viewing things differently: the dimensions of public perceptions of legitimacy. *Criminology*, 51, 103–135.
- Tankebe, J., Reisig, M. D., & Wang, X. (2016). A multidimensional model of police legitimacy: A cross-cultural assessment. *Law and Human Behavior*, 40(1), 11–22.
- Trinkner, R., Jackson, J., & Tyler, T. R. (2018). Bounded authority: expanding “appropriate” police behavior beyond procedural justice. *Law and Human Behavior*, 42(3), 280–293.
- Trinkner, R., Mays, R. D., Cohn, E. S., Van Gundy, K. T., & Rebellon, C. J. (2019). Turning the corner on procedural justice theory: exploring reverse causality with an experimental vignette in a longitudinal survey. *Journal of Experimental Criminology*.
- Tsushima, M., & Hamai, K. (2015). Public cooperation with the police in Japan testing the legitimacy model. *Journal of Contemporary Criminal Justice*, 31(2), 212–228.
- Tyler, T. R., & Jackson, J. (2014). Popular legitimacy and the exercise of legal authority: motivating compliance, cooperation and engagement. *Psychology, Public Policy and Law*, 20, 78–95.
- Tyler, T. R., & Jackson, J. (2013). Future challenges in the study of legitimacy and criminal justice. In J. Tankebe & A. Liebling (Eds.), *Legitimacy and criminal justice: an international exploration* (pp. 83–104). Oxford: Oxford University Press.
- Tyler, T. R. (2006a). Legitimacy and legitimation. *Annual Review of Psychology*, 57, 375–400.
- Tyler, T. R. (2006b). *Why people obey the law*. Princeton: Princeton University Press.
- Tyler, T. R. (2017). Can the police enhance their popular legitimacy through their conduct: using empirical research to inform law. *University of Illinois Law Review*, 1971–2008.
- Tyler, T. R., Jackson, J., & Mentovich, A. (2015). On the consequences of being a target of suspicion: potential pitfalls of proactive police contact. *Journal of Empirical Legal Studies*, 12(4), 602–636.
- Tyler, T. R., & Fagan, J. (2008). Legitimacy and cooperation: Why do people help the police fight crime in their communities. *Ohio St. J. Crim. L.*, 6, 231.
- White, M. D., Mulvey, P., & Dario, L. M. (2016). Arrestees' perceptions of the police: exploring procedural justice, legitimacy, and willingness to cooperate with police across offender types. *Criminal Justice and Behavior*, 43, 343–364.