Response to Dooley and Goodison

RESPONSE TO DOOLEY & GOODISON:

Falsification by Atrophy: The Kuhnian Process of Rejecting Theory in American Criminology

Stephen Farrall and Richard Sparks

(*British Journal of Criminology*, 60, 1: 45-9)

We would like to start our response by thanking Brendan Dooley and Sean Goodison for producing such a thought-provoking paper. We were the (originally anonymous, now self-outed) peer reviewers for this journal. The stimulus and, we should admit, provocation that we received from Dooley and Goodison's paper encouraged us both into producing two of the longest peer reviews either of us can recall writing. So when Sandra Walklate offered us the opportunity to make those reviews the basis for a response to the paper in the *BJCrim* we gladly took up the challenge.

As the reader can deduce both of us were convinced enough of the value of the paper to support its being published. Seeing it now in its published form, we are all the more persuaded that that was the correct judgement. Dooley and Goodison's paper is an unusual one in our observation of recent criminology in directly addressing classical problems in the philosophy of science that are often passed over in silence and perhaps regarded, if only implicitly, as 'old hat'. Moreover, the authors do not merely remind us of the debates between Kuhn and Popper, they want to put the old arguments about paradigmatic succession on one hand and the principle of falsificationism on the other to active use. That is, as well as in their words inviting 'further discourse' (well, here is some!), they want to encourage precision in framing propositions to promote testability wherever possible.

We admire the seriousness of this argument and think that under some conditions, and for some purposes, it is persuasive. We also, however, continue to have a number of worries about and disagreements with aspects of the argument. Some of these are themselves technical and relate to methodological aspects of both the quantitative and qualitative dimensions of the study. Our broader concerns, however, include whether in framing the issues as they do, Dooley and Goodison, consciously or by implication, also take a somewhat restrictive position on what criminology properly so-called is and should be. This goes to a range of questions about the multiplicity of criminology's objects and its disciplinary connections that are raised but not we think fully developed in the paper.

Moreover, the exclusively American[1] focus of the paper brings to the surface a number of differences between dominant conceptions of the criminological enterprise in the United States and elsewhere in the world that are widely acknowledged but rarely discussed (or not at least in the United States) as matters of theoretical import. The American-ness of this paper is in a sense a limitation, but it is also an opening towards a more systematic comparative sociology of criminological knowledge than currently exists. If other traditions or cultures of criminology exist, as they clearly do, this tells us something of note about their conditions of emergence and existence, and this is a stimulus towards other kinds of inquiry (see for example Savelsberg et al., 2015).

[1] There is a slight confusion of terminology in Dooley and Goodison's paper. The title refers to 'US' but much of the text refers to 'American' studies (which could include those in Canada, or for that matter, Southern America). Since neither the studies nor the journals in which they are published are cited, we have referred to 'America/n' throughout.

The paper sets out to explore the extent to which there is an inherent bias in the published work in criminology as it pertains to the testing of theories. The authors argue that, as a field, criminology is not very good at publishing the results of efforts at testing theories that refute the theories being tested. What results, they argue, is that rather than offering many examples of outright falsification, the field shows a tendency towards change through a process that they term ‘atrophy’ (whether by ‘exhaustion’, ‘indolence’ or ‘assault’.) The analysis proceeds in two stages – a quantitative content analysis of around 500 articles in notable American journals, and a set of qualitative interviews with 17 senior American criminologists. Although the authors, and certainly many of their distinguished group of interview respondents, may in principle favour Popperian falsification, in reality the real processes of change in the field over time look rather more like a collective shift in focus or fashion. In other words, as the title of Dooley and Goodison’s paper indicates, real scientific activity tends to proceed more as Kuhn anticipated than as Popper stipulated.

As noted above, and as the paper’s title acknowledges, this is not a study of criminology *per se*, but of American criminology. All of the journals are American. All of the leading criminologists interviewed are American scholars (most are men and white, just for the record). This geographical focus may (or may not) alter the findings but it does speak to an American culture of research design which may look distinctly different from what criminology is and what it represents in other academic cultures or countries. We further suspect that the absence of focus on theory testing and falsification would be *more* pronounced in most other national criminological scenes. This would make it no doubt *even worse* in the view of the majority of Dooley and Goodison’s interviewees. This however would be to depict as a simple deficit what are probably better thought of as embedded differences of historical formation and institutional setting. In other words, it would be beneficial to say more (and this applies within the American case just as much as it does outside it) about what those who do not primarily do theory testing actually do do and not only on what they don’t.

The exclusively American focus of the paper is perhaps related to other issues in the construction of knowledge: the authors rather casually assume that a theory can be tested and either confirmed or rejected. Although this works well on paper, this approach ignores ‘real world science’. In other words, via an iterative process, a theory may be confirmed at time i, but rejected at time ii (since the domain assumptions of the theory no longer apply for some reason). Similarly, a theory may be rejected in Country A, but supported in Country B (since it was developed in Country A and Country A is different to Country B in some key way). Because the authors relied only on American studies, and only selected studies published between 1993 and 2008 (15 years; although the time of the actual fieldwork is not mentioned) these possibilities cannot enter their analyses or thinking. However, this ‘constant cause’ thinking runs counter to what we know from comparative and historical studies (that causes operate differently in different places and at different times).

This is compounded by the consequences of the peer review system employed by journals (as the authors briefly note on the last page); namely that the originators of the theories being tested would be amongst those who would be reviewing journal submissions and hence could find grounds to reject the articles, meaning that the published data (on which they draw) is biased. Scholars who propose theories are often the very same people most likely to be invited to review for journals. The blind peer review system allows for reviewers to reject papers which might embarrass them if they were published. What would be needed would be an examination of the grey literature and additional interviews with a wider pool of scholars (who might be able to provide – or refute - evidence of corruption in the review process). In addition to this, another bias which entered the research design was the reliance on journals as the repository of knowledge. This decision ruled out (for example) Ezell and Cohen’s testing of three life-course theories (2005) and Johnson’s (1979) similar efforts – both of which were books.[2] Perhaps book-length contributions enable authors the space to justify their rejections more than the comparable shorter journal articles do? It is also unclear if the assessments which were reached were simply of the "we studied theory X and rejected it" style, or if these were publications which took several studies and tested and compared them using the same data set (which is the better approach, and what Ezell and Cohen and Johnson did).

Similarly, tests of theories are only as good as the data at hand: any test, irrespective of the outcome (reject, confirm, equivocate) is only as good as the data on which it is based. This may vary between countries and over time (see above) but means that the testing of theories is an iterative process; a theory may be confirmed (or rather, not rejected) for a while only to be rejected when better data becomes available. If this was the case, one might expect that there would be fewer 'rejects' than 'confirms' because as soon as a theory is rejected it 'dies'. Finding that only 12% of the studies were 'rejects' therefore would not indicate failure on the part of the criminological community; rather it would represent science working as planned – we accept something as plausible (or 'correct') *until* it is rejected. An alternative approach to exploring this topic might be to produce a case study of a theory which was rejected, following it's 'life-course' from proposal to testing to rejection and abandonment.

On the topic of theory rejection, it ought to be noted that rejecting a theory may rightly involve a higher standard of research design: if the process we outlined above *is* what happens, then it may also be the case that in order for a theory to be rejected, the 'reject' study needs to be of the highest possible quality (and few studies really *are* that well-designed or executed!). Hence, again, a low level of rejection might be expected – since, no one wants to reject a theory which might actually be valid on the basis of one or two errant studies. In effect, for a theory to be rejected, one needs several studies, but each study a) has to be of the highest quality and b) has to individually negotiate the blind review processes which means that it can more easily be rejected on the basis of either good, scientific caution or bad, game-playing reviewers (on which, again, see above). Finally, more thought needs to be given to the development of a more nuanced view of confirm/reject assessments. It is not clear exactly what is meant by reject or confirm. Is this a total rejection of *all* assumptions or a rejection of *some* causal explanations which suggests possible refinement rather than dismissal? Some theories may sound 'rejected' because a step in the explanation is not verified, whilst the rest of it is supported. Is this a reject or a neutral?

The most significant effect of the timing, location and focus of the study lies in its implicit construction of what criminology primarily is. All the theories discussed in the content analysis are theories of *offending.* Although the authors are quite careful in acknowledging criminology's multi-disciplinary character and the existence of its critical traditions, almost all the work they discuss (and that of by far the majority of that of their interviewees) have this in common. They are theories of offending rather than of victimisation, for example, or crime rates, or criminal justice processes or institutions, still less histories of crime or studies of popular culture or technological change. It may be that this has biased the results in some way, and that, had theories which had been developed to account for a different phenomenon been selected the conclusion may have been different.

Why is this a problem? The authors are quite entitled to study and discuss the relevance of falsification in relation to a certain important body of work, whether or not it constitutes the whole of the criminological enterprise. The simple solution, faced with such decisions, might be to say:

[2] Whilst some of these points are acknowledged by the authors, simply acknowledging a weakness in a research design does not mitigate against it.

those propositions that are capable of being framed in a falsifiable manner (because they are making an empirical claim about what is likely to occur under strictly specified conditions) should be so. Others may answer to different criteria of evaluation and conventions of rigour. So which are these topics; and which kinds of work can feasibly aspire to be like that? This would still leave a considerable space for historical, comparative, interpretive, ethnographic, critical, speculative, normative and other modes of work that are unlikely in principle to be amenable to those strictures. This seems to us to be potentially both more precise and more accommodating of plurality than what is disclosed by the authors' approaches here.

We see here an opportunity to move from a discussion of the apparent problem (if such it be) of the relative infrequency of outright falsification in criminology to an account of the internal organization of a distinctly plural field. The paper seems to us to lead in this direction but not to explore it very far. As we have already noted a comparative dimension might have helped considerably in this regard. Similarly, a more pronounced historical sensibility, broaching the question of why certain theoretical position gain credence and influence when and where they do, would help locate the relevant debates (their persistence as well as their atrophy) more precisely. Dario Melossi's *Controlling Crime, Controlling Society* (2008) is a work of exactly this kind.

Ultimately, criminology's interdisciplinarity, theoretical patchwork and methodological messiness seem to us to be here to stay, and if anything likely to get more pronounced, whatever some of the august gentlemen interviewed by Dooley and Goodison may wish. The ceaseless mutations of crime and control, their endless dialectic with new technological formations, their chronic involvement with politics, render the field uncontrollably plural. We had all better get used to the idea that there are a lot of people on the criminological camp-site and they are not all equally interested in the same things. Why would that be? Presumably – as the paper also seems to acknowledge at other points – because the topics themselves that the field exists to explore are many-sided and call on a wide variety of skills, kinds of training and forms of expertise.

This seems to us to be the implication of the analysis, and to be present in several of the statements of the august subjects of the qualitative study, and to be implicated in some of the diverse forms of 'atrophy' that the authors interestingly propose. So when Tittle or others say that we keep coming back to questions of social change they are acknowledging (surely?) that in intellectual and disciplinary terms (in terms that is of the kinds of conceptual resources we need in order to deal with it) this goes quite some way away from testing propositions about police tactics at hot spots. If we are to have a serious conversation about falsifiability then we also need to have a serious discussion about levels of explanation, but that is a topic raised but not really explored here. Either criminology is about one thing (falsifiable propositions about the explanation of crime), or it's about many things, not all of which can or should take that logical form.

<u>References</u>

Ezell, M. and Cohen, L. (2005) *Desisting from Crime*. Oxford: Oxford University Press

Johnson, R. (1979) *Juvenile Delinquency and Its Origins*. Cambridge: Cambridge University Press.

Melossi, D. (2008) *Controlling Society, Controlling Crime*. Cambridge: Polity Press

Savelsberg, J., Hughes, L. A., Kivivuori, J., Short, J. F., Sozzo, M. & Sparks, R., (2015) 'Criminology, History of', *International Encyclopedia of Social and Behavioral Sciences.* (2nd edition) Amsterdam: Elsevier, (pp238-243)