



New South Wales

Special Commission of Inquiry into LGBTIQ hate crimes

Expert Report Associate Professor Austin Lovegrove

A. Introduction

1. I acknowledge that I have read the Expert Witness Code of Conduct in Schedule 7 to the Uniform Civil Procedure Rules 2005 (NSW) and agree to be bound by it.
2. My name is Austin Lovegrove. I am an Associate Professor and a Principal Fellow in the Law School at the University of Melbourne. I was head of Criminology at the University of Melbourne from 1985-1987 and from 2002-2005. I have no specialisation in issues connected to gay hate, but I do have significant expertise in the acquisition, analysis and presentation of empirical data in the context of the social and behavioural sciences.
3. My expertise in those areas arises from my experience in:
 - (1) Academic training;
 - (2) Teaching; and
 - (3) Personal research.

Academic Training

4. My undergraduate studies were in the Department of Psychology, University of Melbourne. During each of the four years of my honours degree, I completed one subject on research methodology, research design, research techniques and the statistical analysis of empirical data. A minor research thesis was part of both the third and fourth years, which required students to plan and conduct an empirical study. Both studies involved studying the development of young children's thought processes.
5. I continued my studies in the Department and completed a Masters by research in around 1970-1971. The aim of my Masters was to identify differences with respect to personality traits and attitudes between young teenage criminal offenders and non-offenders. To this end, I developed and applied a psychometric test using a sophisticated statistical technique, namely, a two-stage principal components factor analysis.
6. From around 1972-1975 I completed a PhD in the Psychology Department, University of Melbourne. My PhD involved an analysis of the decisions individuals make whilst they are driving vehicles. The aim of my project was to identify the environmental cues drivers rely on to determine whether they are travelling at a safe speed as they approach a certain type of intersection. The study involved the experimental manipulation of variables in real-world settings, and was able to explain why a significant percentage of drivers approach this type of

intersection at unsafe speeds. My PhD thesis was examined by a leading academic human factors engineer and a behavioural psychologist.

Teaching

7. Apart from teaching a general criminology course in my early years, I taught three subjects: Research Methods in Criminal Justice; Psychology, Crime, and the Criminal Justice System; and Sentencing. These were taught at the third-year undergraduate level and the graduate level.
8. Research Methods in Criminal Justice introduced students to varieties of research design and a range of common research techniques for data collection (for example, rating scales, questionnaires, structured and semi-structured interview schedules) in addition to sampling strategies. In my classes, students were also exposed to published research to provide them with clear illustrations of relevant designs and techniques. For the purposes of assessment, each student was required to plan a research study.
9. Psychology, Crime and the Criminal Justice System had a large empirical component (50%). Here, students were introduced to published empirical studies covering a variety of research designs and techniques. The studies also examined various components of the criminal justice system (including causation, policing, courts, corrections).
10. The Sentencing course covered the following topics: Legislation; Proportionality and just deserts; Incapacitation; Rehabilitation; Deterrence; Seriousness and severity; Victim; Offender; Judge; Public opinion; Assessing the effectiveness of sanctions; Controversies in sentencing; Guidance for sentencers; and Sentencing research.

Research

11. I have conducted behavioural and social research across multiple research problems and drawing upon a variety of methodologies; the one common feature is that all have involved real-world settings and applied problems. My research is represented in the most prestigious international publications. A list of my research in those publications is Annexure A to this report.
12. Since completing my PhD, I have conducted research as an academic member of staff in the Criminology Department at the University of Melbourne. One major line of research aimed at identifying how judges make sentencing decisions. This research included two specific projects. Both attempted to develop decision models for cases of burglary determined in the higher courts. One study focused on single cases comprising one count; the other dealt with the

problem of cases involving multiple offences where, a sentence having been determined for each of the comprising counts, the problem arises as to what is the proportionate and appropriate sentence for the case in its totality. In the first study, I conducted a decision analysis of judicial determinations and sentence for a set of fictitious cases. Following a Multi-Attribute Utility Measurement (MAUM) technique, I considered how judges categorised various elements of the offence of burglary (value of theft, degree of organisation, degree of violence) on scales of less serious to more serious, assessed the judges' importance weighting for each element, and then combined these 'scores' mathematically. The approach was quantitative. For the second study, a qualitative 'thinking aloud' technique (a variation of protocol analysis) was used. Here, individual judges gave an account of their thinking as they determined the sentence for a range of skeleton cases which were formulated to collectively represent various challenges in sentencing cases involving multiple offences. The approach was qualitative.

13. A third major research project addressed the perceived problem of judges' sentencing decisions being too lenient. The common explanation for this is that judges are out of touch with the public's feelings about appropriate punishments for various types of crime. To this end, four experienced sentencing judges each presented one of their cases to eight small groups selected from around the state. The 32 groups comprised a total of 470 people drawn from the community. These individuals then individually and anonymously determined what they thought to be the appropriate sentence in the case and identified the critical factors in the decision. Their answers were recorded by way of a semi-structured questionnaire. This study was regarded as an innovative methodological approach to this problem.
14. My PhD and these three studies have been accepted at the highest levels of international scholarship.
15. An article reporting the key experimental study of my PhD ('Approach speeds at uncontrolled intersections ...') was accepted by the *American Psychological Society* for publication in its premier journal for applied research, the prestigious *Journal of Applied Psychology*.¹
16. My research into judicial decision making was accepted for publication in two prestigious research monograph series in criminology. These are *Judicial Decision Making* (Springer, New York) and *The Framework of Judicial Sentencing* (Cambridge University Press, Cambridge

¹ Professor Austin Lovegrove, 'Approach speeds at uncontrolled intersections with restricted sight distances' (1978) 63(5) *Journal of Applied Psychology*, 635.

England). For both series, the editor was a professor of criminology at Cambridge University, and recognised internationally as one of the most eminent empirical psychologists/criminologists.

17. The research on the public's sense of justice in sentencing in regard to judges' sentencing decisions has also been published in prestigious academic outlets. 'The Pernicious Impact of Perceived Public Opinion ...' was included as a chapter for the edited book, *Mitigation and Aggravation at Sentencing*.² The editor was professor of Criminology at Oxford and recognised internationally as one of the most eminent empirical psychologists/criminologists in the field of sentencing research. Most of the academics contributing to this book enjoyed international recognition.
18. Publications which described this study appeared in academically significant international journals too. A general report of the study ('Putting the offender back into sentencing: An empirical study ...') was published in the leading journal, *Criminology and Criminal Justice*, the journal of the British Criminological Society.³
19. The results of this research and several of the author's observations were quoted with approval in a reported judgment of the Victorian Court of Appeal (*WCB v R* (2010) 29 VR 483). An article reporting an aspect of the study – the application the data to a legal theory – was accepted for publication in the prestigious *Cambridge Law Journal* (the Cambridge Law School's journal).⁴ The then-President of the Victorian Court of Appeal subsequently sought to have the Government fund an extension of this study.
20. One other publication is relevant to the breadth and quality of my research output, namely, 'The Listing of Criminal Cases in the Crown Court as an Administrative Discretion'.⁵ This involved interviews with officials responsible for the listing of criminal cases at Crown courts, and aimed at understanding what, if any, consideration was given to the judge hearing the case. The editor, the Vinerian Professor of Law at Oxford University, an eminent criminal lawyer and

² Professor Austin Lovegrove, 'The Pernicious Impact of Perceived Public Opinion on Sentencing' in Julian V Roberts (ed), *Mitigation and Aggravation at Sentencing* (Cambridge University Press, 2011) 188.

³ Professor Austin Lovegrove, 'Putting the offender back into sentencing: An empirical study of the public's understanding of personal mitigation' (2011) 11(1) *Criminology & Criminal Justice* 37.

⁴ Professor Austin Lovegrove, 'Proportionality Theory, Personal Mitigation, and the People's Sense of Justice' (2010) 69(2) *Cambridge Law Journal* 321.

⁵ Professor Austin Lovegrove, 'The Listing of Criminal Cases in the Crown Court as an Administrative Discretion' (1984) *Criminal Law Review* 738.

criminologist, not only accepted it for publication but, believing this was high-quality research and raised a matter of importance, sent as copy to Editor of *The Times*, where it was reported with some prominence.⁶

21. In summary, I consider that my training, teaching and research have equipped me to evaluate methodologically the two research studies which are the subject of my report.

B. Documents

22. I have been provided with the following documents:

- (1) A letter of instruction from Mr Enzo Camporeale dated 11 October 2022 (annexed to this report and marked "B");
- (2) The Inquiry's Terms of Reference;
- (3) The Final Report of Strike Force Parrabell (**Parrabell Report**);
- (4) The Coordinating Instructions for Strike Force Parrabell (**Coordinating Instructions**);
- (5) A joint written submission, dated 5 November 2018, by Professor Willem de Lint and Associate Professor Derek Dalton of Flinders University (the two principal members of the academic review team for Strike Force Parrabell) to the Standing Committee of the NSW Legislative Council;
- (6) The transcript of the joint oral testimony of Professor de Lint and Associate Professor Dalton to that Committee on 28 November 2018;
- (7) An academic article by Professor de Lint and Associate Professor Dalton entitled 'Anatomy of Moral Panic: The "List of 88" and Runaway Constructionism', (2021) 29 *Critical Criminology* 723 (published online on 31 July 2020); and
- (8) A document authored by McLaughlin et al entitled "Responding to Hate Crime: A Multidisciplinary Curriculum for Law Enforcement and Victim Assistance Professionals", published by the US National Centre for Hate Crime Prevention, Education Development Centre (2000).

23. Where I have had regard to, or relied upon, any other documents for the purposes of this report, this is clearly noted in the report.

⁶ Frances Gibb, 'Court Officers' "bias" in listing cases', *The Times* (London, 12 December 1984) 2.

C. Assumptions

24. For the purposes of preparing this report, I have been asked to make the following assumptions:
- (1) In or about 2015, the NSW Police Force (**NSWPF**) established Strike Force Parrabell.
 - (2) Strike Force Parrabell was a successor to “Operation Parrabell”, which had commenced in about 2013.
 - (3) Strike Force Parrabell was established to review a list of 88 previously investigated deaths of persons, between 1976 and 2000, to determine if a “sexuality or gender bias” was a contributing factor”: see Coordinating Instructions, p 2.
 - (4) Elsewhere in the Coordinating Instructions (at pp 2 and 3) and in the Parrabell Report itself (e.g. at pp 17-20), other terms were used besides “sexuality or gender bias”, including “gay hate crimes”, “bias crime” and “anti-gay bias”, as also indicating the objective or purpose of the Strike Force.
 - (5) Strike Force Parrabell was to review matters that had already been investigated by the NSWPF. It was not to, and did not, reinvestigate any matters. It did not contact any witnesses, suspects or family members. Rather, it reviewed written “holdings” from NSWPF files. Its aim was to arrive at a determination as to whether any of the 88 deaths were in fact motivated by an “anti-gay bias” (or the like), rather than identifying and prosecuting offenders: Coordinating Instructions pp 3, 14; Parrabell Report pp 19-22.
 - (6) Various officers from the NSWPF were seconded to Strike Force Parrabell over an 18-month period: Parrabell Report, p 20. The names and ranks of those officers are at p 6 of the Parrabell Report.
 - (7) NSWPF engaged academics from Flinders University (**the academic team**) to conduct a review of the work of the Strike Force: see Parrabell Report at p 21.
 - (8) Pages 1 – 46 of the Parrabell Report comprise the report of the Strike Force officers, while the longer part of the Parrabell Report, at pages 47 – 133, consists of the academic review by the Flinders University academic team.
 - (9) In conducting their review of the cases, Strike Force Parrabell officers used a “Bias Crime Indicators Form” (**BCIF**), said to be used by the NSWPF Bias Crime Unit: Coordinating instructions pp 3-4; Parrabell report pp 20-21.

- (10) That BCIF Form is set out in full at pp 4-13 of the Coordinating Instructions.
- (11) According to the Coordinating instructions (pp 3-4 and footnote 1), and the Parrabell Report itself (see pp 67-70 within the academic review section of the Report, and footnote 20 thereto):
- (a) Indicators 1-9 in the BCIF had been derived from the US document authored in 2000 by McLaughlin et al (document 7 above), while
 - (b) Indicator 10 (as to “Level of Violence”) had been developed by the NSWPF Bias Crime Unit.
- (12) The methodology employed by the Strike Force officers, including their use of the BCIF, was set out at pp 3-4 of the Coordinating Instructions and at pp 19-22 of the Parrabell Report. It was the subject of comment by the academic team at pp 65-70 of the Parrabell Report.
- (13) In respect of each of the cases considered by Strike Force Parrabell:
- (a) the BCIF form was completed by one or more of the police officers who comprised the Strike Force, in the sense that the 10 sets of “prompts” and “indicators” in the left hand column of the form were answered in writing in the right hand column;
 - (b) such police officers were of varying ranks and experience;
 - (c) such answers, and thus such completed BCIF forms, were of varying lengths and composition;
 - (d) the answers to such “prompts” and “indicators”, in each particular case, were composed by whichever Parrabell officer or officers was or were assigned to that case, derived from the views formed by such officer/s in the light of their reading of whatever historical files were available in that case; and
 - (e) in all or most completed forms, there was a final box headed “Summary of Findings”, in which the assigned Parrabell officer/s ascribed one of four possible descriptors to the case, namely either “Evidence of bias crime”, or “Suspected bias crime”, or “No evidence of bias crime”, or “Insufficient information”.
- (14) In Recommendation 3, the Strike Force officers expressed the view that “the current system with 10 bias crime indicators” was “not user friendly for operational police”:

Parrabell Report, p 39.

- (15) The academic team expressed reservations about the BCIF tool used by the NSWPF: Parrabell report pp 68-71 and footnote 20.
- (16) The academic team used a different methodology, as outlined by them at pp 56-58, 70-71 and 79-91 of the Parrabell Report.
- (17) The academic team created and applied their own definition of “bias”: Parrabell Report, pp 82-83.
- (18) The academic team then categorised bias crimes based on whether they were:
- (a) “proactive” or “reactive”, and
 - (b) “associative” or “non-associative”: Parrabell Report, pp 88-90.
- (19) On 5 November 2018, Professors De Lint and Dalton made a joint written submission (document 4 above) to the Standing Committee of the NSW Legislative Council on Social Issues, which was conducting an Inquiry into Gay and Transgender Hate Crimes between 1970 and 2010 (the Parliamentary Committee), in which they said, among other things:
- “There are difficult questions to be asked when assessing whether an incident is a bias crime and quite often investigators do not have the information to answer them” – at [3.4];
 - “The tool used by the NSWPF, the Bias Crime Indicators Review Form (BCIRF), draws its ten indicators from the National Centre for Hate Crime Prevention ... It leaves too much discretion and does not rank or prioritise among the indicators and is being discontinued by the NSWPF” – at [3.6]; and
 - “[W]e designed a simple three-part test of bias is as follows. Bias crime: expresses a categorical animus (directed at a person or group on the basis of his/her perceived identification with a vulnerable group); produces an act that intentionally, by way of criminal predation on the basis of that categorical animus, causes harm to that person or group; is mitigated or aggravated by an offender’s contemporaneous associations that are linked by a commitment of denunciatory non-identification with the vulnerable person or group” – at [3.8].
- (20) On 28 November 2018, Professors De Lint and Dalton gave joint oral testimony (a

transcript of which is document 5 above) to the Parliamentary Committee, in the course of which they made statements to the following effect:

- At p 11 – that the material with which the academic team was provided, and on which they based their analysis, consisted of “case summaries” (which the Inquiry at this stage, subject to any later clarification, understands to be a reference to the completed BCIF questionnaires with the ten sets of “prompts” and “indicators”) created by the Parrabell police officers (in 2015-16);
- At pp 11-12 and p 14 – that the academic team did not look at any of the primary documents which “sat behind” those summaries;
- At pp 10, 14 – that the academic team looked at the case summaries “to determine whether or not the attribution of bias that was provided (by the Parrabell officers) made sense to us”;
- At p 15 – that the police did not attempt to ascertain whether other homicides in the relevant period, beyond the list of 88, might be gay bias related; and
- At pp 17-18 – that the academic team treated cases of anti-paedophile bias (where in their view that was indicated, from the case summaries) as separate and distinct from anti-gay bias.

D. Questions addressed in this report

25. Having made the assumptions listed above in section C, I have been asked by the Commissioner to address the following questions:

- (1) What is your view as to the appropriateness of the methodology used by the Strike Force Parrabell officers, including the use of the BCIF, having regard to the objectives of the Strike Force (as to which see assumptions 3, 4 and 5 above)?
- (2) What is your view as to the appropriateness of the different methodology used by the academic team, having regard to those objectives?

E. List of Acronyms used in this report

26. In this report, I use the following acronyms:

- (1) Parrabell Report – Police: PR;
- (2) Parrabell Report – Academic: AR;

(3) Co-ordinating Instructions to Strike Force Parrabell: CI.

F. Question 1: What is your view as to the appropriateness of the methodology used by the Strike Force Parrabell officers, including the use of the BCIF, having regard to the objectives of the Strike Force (as to which see assumptions 3, 4 and 5 above)?

Assessment of methodology used by Strike Force Parrabell

27. My assessment of the Strike Force's research methodology follows a logical sequence and is set out in three parts. First, background matters are examined, including the objectives of Strike Force Parrabell, research design, definitions of 'hate' and 'gay' as adopted by the Strike Force, and the reasons for selection of the BCIF as the instrument for assessing whether gay hate was a significant motive in each of the 88 cases of suspected homicide. Second, I assess how the BCIF was used, having regard to its actual use by the Strike Force researchers as instrument for identifying gay-hate crime. Third, I examine the Strike Force researchers' numerical analysis of their data, having regard to their four-part classification of the incidence of gay-hate crime.

Background Matters

Objectives of Strike Force Parrabell

28. The principal objective of the Strike Force (variously stated), was to investigate the incidence of 'gay-hate' as a significant underlying 'causal' factor in the 88 deaths on the Strike Force Parrabell 'list' (the potential significance of this definitional variety is addressed below).⁷
29. The secondary objectives (again, variously stated) included bringing the NSW Police Force and the Gay community closer together;⁸ addressing the significant angst in the LGBTIQ community regarding police investigative propriety;⁹ addressing continued unrest in the LGBTIQ community regarding bias motivation, especially related to the 88 deaths on the 'list';¹⁰ and showing proactivity on these matters.¹¹

Research Design

30. Before proceeding to the detailed evaluation of the Police Report and of the Academic Report, I set out a necessarily brief overview of:

⁷ NSW Police Force, *Strike Force Parrabell Final Report* (Report, 2018), 19-20.

⁸ Ibid 18.

⁹ Ibid 14.

¹⁰ Ibid 18.

¹¹ Ibid 14.

- (1) the terms “reliability” and “validity”, what they mean (and by implication why they are important), and the means by which they should be assessed; and
- (2) the concept of causation, and how the method adopted to establish any relationship between two variables determines how confident one can be that the relationship represented is true.

Both issues are central in each report to understanding the measurement of hate and its role in the behaviour in question.

Understanding reliability and validity and their assessment

31. Reliability is a term used in behavioural and social science to describe instruments and assessment tools. An instrument is reliable if two individuals using the instrument would independently reach the same judgments about the thing which is being measured by the instrument.
32. Validity is a term used in behavioural and social science to describe whether an instrument or assessment tool measures what it purports to measure. In this case, an instrument would be more valid if it were more effective at measuring gay hate and would be less valid if it measured something else.
33. There are a variety of numerical and mathematical tools used to measure reliability and validity (percent agreement is the measure most readily understood). Poor reliability necessarily reduces validity, but there may be high reliability yet low validity (the individuals’ judgements are consistent but erroneously based or based on different criteria).
34. The Police Parrabell study used a behavioural instrument (the BCIF) to determine whether hate was involved in any of the suspected homicides in any of the of 88 cases. So too the academic study. In fact, the instrument used in the academic study is better described as an instrument comprising two components: a definitional component defining hate and a classificatory component for the purpose of assigning cases to categories according to the circumstances in which the hate was expressed and to the nature of the hate.
35. In view of each instrument’s centrality in the relevant study, the question arises as to its fitness for purpose, having particular regard to its reliability (in this case, whether separate individuals would independently agree that the case involved a hate crime) and its validity (in this case, whether those assessments, even if made reliably, were in fact identifying hate crimes).
36. If an existing instrument or assessment tool is being used, there must be evidence of its

reliability and validity from past research, in order for the exercise to be methodologically sound.¹² It would then be a matter of whether this research, which may have been conducted in another country, could be reasonably assumed to apply in the present circumstances.

37. Where an instrument is newly developed and the judgment it is used for is an end in itself (here, identifying the presence of hate in criminal cases), as in the academic study, then the researchers must first test the reliability and validity of the instrument they have developed for use in the study.
38. The process of testing the reliability and validity of an instrument would involve the following steps.
39. First, the researcher would establish two criterion groups, that is: (1) one comprising cases that involve gay hate; and (2) one comprising cases that do not involve gay hate (in fact, there may be one or more additional groups, for example: (3) cases that involve 'suspected gay hate')¹³.
40. The researcher would then approach one or more people with high-level experience in the assessment of criminal evidence or investigation of criminal circumstances (e.g. senior judges) who are in a position to categorise a particular universe of cases into the criterion groups. Such people will necessarily have, from training and long experience, the ability to make such assessments (what might be described metaphorically as a 'Rumpolian nose'). It may also be appropriate for this expert person or group to ensure that they conduct the categorisation apprised of what the gay community would regard as the signs of gay hate in the circumstances of these types of crime, as they were manifest in the era and milieu which are the subject of the investigation. I note that this does not necessarily mean that the observations of the gay community would be accepted uncritically in respect of this (since to do so might otherwise result in the over identification of gay hate). Rather, this input would be approached with the attitude that the gay community may be aware of manifestations of gay hate of which others in the community are not aware.
41. Second, the researcher would then develop an instrument that attempts to identify gay hate in the circumstances of a crime. In association with this, it might be decided to use a psychological test intended to identify gay-hate attitudes in individuals. The instrument measuring gay hate

¹² I note that the BCIF, in the form used by Strike Force Parrabell, was not an existing instrument or tool.

¹³ This process is not necessary in all circumstances; for example, the use of a newly developed instrument to test an academic theory.

may take various forms, and levels of complexity.

42. At the highest level of complexity, the instrument may take the form of a decision framework. It would reflect the way in which users of the instrument (say, experienced detectives) decide whether a crime involves gay hate. To this end, investigators may be asked to assist with the development of the instrument by assessing a set of real-life cases. The decision framework would attempt to identify the potentially relevant factors and to specify how this information is to be combined as a global judgement. There is a range of formal techniques a researcher can use to facilitate this. The instrument would appear as a largely qualitative guideline.
43. At the intermediate level of complexity, the instrument may take the form of a psychometric scale; the instrument would identify the major factors in the judgement. Potentially relevant factors may be informed by the communicated experience of senior detectives, and those who have been the subject of gay hate, and perhaps the relevant academic literature. Individual judgements made in respect of these factors for a set of cases would then be combined, this combination generally being made additively and with equal weighting. Thus, the researcher would determine which of the (say) ten to twenty or more factors apply to a case, assign a score of one to each of them, and add the scores. It follows that the higher the score, the more certain it is taken to be that hate is involved in the case. It would appear as a largely quantitative guideline.
44. At the lowest level of complexity, the instrument would take the form of a checklist. It would comprise no more than a list of pertinent clues.
45. At these lower levels, the instrument would not say anything about how the judgement should actually be made, and would almost certainly greatly oversimplify the actual decision making. A more complex instrument has the potential for higher validity but will likely be undermined by lower reliability. A less complex instrument has the potential for high reliability but the price of simplicity will likely be low validity.
46. There is no set methodology for developing an instrument; it will depend in part on the nature of the data and its intended use.
47. Third, the researcher would select a sample representing the individuals for whose operational use the instrument is intended (e.g. junior detectives).
48. Fourth, the researcher would present the criterion cases, unidentified by category, to this group of individuals (here, junior detectives), and have them independently use the instrument

to identify which of the cases involved gay hate.

49. Fifth, the researcher would measure the inter-rater reliability of these assessments (the levels of agreement between the raters across the cases (see above)). A high level of agreement is required of a sound instrument. The greater the complexity of the judgement and the more discretion which falls to the individual, the lower will be the reliability of the process; moreover, complexity will soon exceed an inexperienced individual's capacity to form a coherent holistic judgement.
50. Finally, the researcher would compare the extent of agreement (percentage agreement) between (say) the junior detectives' assessments of the presence or absence of hate with the cases involving gay hate as defined by the expert individual or group referred to above in [40]. Where a multi-faceted concept such as hate in manifest circumstances is being measured, achieving more than a modest to good level of validity would be difficult for a number of reasons. In this case, it is possible that individual assessors may be uncertain as to what the concept of "hate" means and how it is individually manifest in the circumstances of a range of actual cases. Some will apply their own understandings – with these differing across raters – or muddle along doing their best where the definitions of the concepts applied are inadequate. A high level of agreement between assessment and criterion is required of a sound instrument.
51. Neither the police, in adopting the BCIF, nor the academic team, in adopting their own framework, appeared to test the reliability and validity of the instruments they applied to identify hate. In my view, neither instrument can be regarded, in view of the absence of appropriate evidence, as fit instruments for identifying gay-hate crimes in the list of 88 cases.
52. Moreover, it follows from what I have said about the foundational role of expert groups in the validation of any behavioural instrument and the professional skills required of the persons who categorise the cases according to criterion groups, that I, as an academic, would not think of myself as having the criminal knowledge or investigative skills to determine gay hate as a motive in the circumstances of violent offending in general or in specific cases.

The meaning of 'cause'

53. Implicit in the Police and academic research studies attempting to identify the involvement of gay hate in the 88 deaths is the common-sense understanding of gay hate being a causal/significant causal factor. In fact, 'causation' is hard to establish using real-world, as opposed to experimental, data.

54. In the experimental approach, a causal effect of (say) a drug is in principle easily established by administering the drug to one of two randomly chosen groups and looking for the effect of the drug in that group.¹⁴ However, trying to establish causation in the social world is problematic, because the lives of individuals cannot be manipulated, and the effect of a factor as a motive cannot readily be observed. A quasi-experimental approach is often used to attempt to uncover a cause.
55. By way of example, consider a supposed relationship between poor home discipline and subsequent criminal behaviour. In this approach, families where there is good discipline will be selected and formed into one group and families where there is poor discipline will be selected for the comparison group. Then the incidence of criminal behaviour in each of the groups is measured. However, there will be uncertainty whether any identified association between poor discipline and offending is truly causal or merely correlational. It might be that the real cause lies in poverty; that poverty causes the poor discipline and independently causes the offending. This third factor (in this example, poverty) is referred to as a confounding variable, because it influences both the independent variable (the cause which is being measured) and the dependent variable (the effect which is being measured). In this example, the alternative explanation or this effect of the confounding variable can be averted by matching both groups on economic circumstances. Yet what about other possible explanatory differences between the two groups? There are practical limitations to the number of matched variables, and there is the real possibility of an unsuspected causal factor being left unmatched. If there is an unsuspected causal factor which has not been matched, the conclusions of the research will be invalidated.
56. This type of research design aims to identify causal links, that is, whether a certain type and level of hate causes violent behaviour. These studies identify general principles about behaviour. They are not concerned with the operation of a factor in a single specific situation (say, whether hate was causal in a particular homicide).
57. Yet their relevance here is to bring caution to our interpretation of the conclusions drawn by the Police and academic team based on the instruments they employed to identify gay hate. Their instruments represent a more or less formal means of identifying the motive of hate in

¹⁴ In fact, additional controls are required when humans are involved. Thus, for example, a placebo group is required to remove the possibility of an imagined effect on the part of the person receiving the drug.

the circumstances of a crime, by way of the presence of certain factors being taken to indicate hate and the absence of these factors being taken to indicate the absence of hate. This is the detection of cause by association. This association represents what might be called evidentiary causation to distinguish it from experimentally established causation. This is because evidence of hate as causal is imputed from facts found in the circumstances of the crime, facts interpreted as the manifestation of a causal link, with hard evidence of the actual causative link between the motive and behaviour being unavailable. Thus, it represents a plausible cause or part cause; it will come with greater or lesser certainty, depending on the nature of the evidence.

58. In respect of section 21A(2)(h) of the *NSW Crimes (Sentencing Procedures) Act*,¹⁵ the notion of cause behind the words ‘motivated by hate’ is necessarily interpreted as what is called here evidentiary cause. This is distinct from experimentally established causation. It ought to be emphasised that the discussion relates to the ‘why’ of a crime, not the ‘who’ or ‘how’.
59. It must also be stressed that ‘causation’ understood in this way does not, of itself, diminish the two studies – indeed, it could not be otherwise, because the studies were dealing with individual cases. I set out the above information simply to offer an enlarged understanding of causation and its detection empirically. Nor does my discussion on causation, of itself, diminish the validity of a causal effect by criminal investigation.

Definitions

60. A problem arises with the Strike Force’s consideration of the terms ‘hate’ and ‘gay’ in their Report. Here the reader is offered something little more specific than what might be called a common-sense understanding of these terms.
61. I consider that the Strike Force researchers failed to engage in (or even expressly touch upon) a theoretical discussion of what they understood to be the essence of ‘hate’ in a crime motivated by hate. Though it is none the worse for the absence of a theoretical discussion, surely a definition of hate is necessary to make more certain a common understanding between researcher and reader? The closest the reader gets to the Strike Force’s understanding of hate

¹⁵ Section 21A(2)(h) provides that it is an aggravating factor to be taken into account in determining the appropriate sentence for an offence if “the offence was motivated by hatred for or prejudice against a group of people to which the offender believed the victim belonged (such as people of a particular religion, racial or ethnic origin, language, sexual orientation or age, or having a particular disability’ (emphasis added).

is that it involves a person's 'sexuality and gender'.¹⁶ In fact, the Academic Report reveals that around the time of the Police study, the NSW Police Force did have a definition of 'bias crime'. Of significance, it states that such a crime may be motivated in whole or in part by bias. The Police may well have applied this understanding in their study.

62. In relation to the use of 'gay' in reference to gay hate, that term is used exclusively throughout the Report. I presume that the Strike Force saw this as encompassing all of the victims. In fact, the victims also include transgender women.¹⁷ For the purposes of research design, it would be more appropriate to define important concepts such as 'gay' with more clarity. For the purposes of this report, I use the words 'gay' and 'gay hate', although I recognise that this does not capture all of the victims.

Selection of the BCIF

63. It is not clear why the Strike Force decided to use the BCIF. It may have been because it seemed plausibly applicable and because nine of the ten indicators found within it were sourced from a reputable Law Enforcement Agency (United States Department of Justice, Office for Victims of Crime).¹⁸ This surely represents tenuous reasoning, far from guaranteeing an instrument fit for purpose. What was needed was independent evidence demonstrating that the BCIF was both reliable and valid.
64. Reliability requires, for example, two independent assessors arriving at the same judgements when applying a measure to the same set of data; the measure of reliability is numerical. No such data are offered as justification for the use of the BCIF here.
65. Validity means that the instrument does what it purports to do (in this case classifies) according to an understood and agreed standard. It would be established, in this case, by demonstrating that the BCIF had previously been used to discriminate between cases deemed by (say) an expert panel to: (i) involve; and (ii) not involve gay hate in a manner that was consistent with panel's judgements. Again, no such data are proffered. Even if the BCIF has validity in respect of the purpose for which nine of the ten indicators were developed in the US, there would be good reason to question its validity here. The BCIF appears to be directed at hate crime

¹⁶ NSW Police Force, *Strike Force Parrabell Coordinating Instructions*, 2.

¹⁷ Indeed, in the discussion of nuances by the Academic team, reference is made to one of the 85 cross-dressing: *Strike Force Parrabell Final Report* (n 7), 105.

¹⁸ *Strike Force Parrabell Coordinating Instructions* (n 16) 3.

generally, not specifically to 'gay' hate. Moreover, there might well be significant cultural and sub-cultural differences in the manifestation of 'gay' hate, considered generally for the US and specifically for Sydney, Australia.

66. Questions about validity also arise due to the nature and content of the BCIF. The BCIF includes a set of ten indicators under each of which there are prompts for evidence (i.e. clues) of gay bias in the circumstances of cases.¹⁹ The US source document lists nine indicators.²⁰ The Strike Force added a tenth indicator, though no explanation was given for its relevance, save that it was based on "research and cases".²¹ These ten indicators were taken by the Strike Force team to cover a range of factual aspects of a crime putatively indicative of 'gay' hate as a significant animating factor. Purpose informs validity. Here, as stated above, one of the objectives of the Strike Force's research is to engage and assuage an alienated and agitated 'gay' community. In my view, the community's understanding of 'gay' crime may bear on the appropriateness of the ten indicators and their specific manifestations in a case. I consider that the opinions and understandings of the gay community may have been profitably obtained with a view to improving the validity of the instrument, at least having regard to the context of the present study.
67. In light of the lack of data on the BCIF's reliability, the reader has no idea whether another group of researchers applying it would classify these same cases as they are classified by the Strike Force researchers. (Indeed, the Strike Force researchers' description of the process they adopted to classify each of the 88 cases suggests reliability will be problematic for the reasons stated below.)
68. Moreover, absent any data on validity, the reader has no idea as to the number of cases classified falsely as involving bias or falsely as not involving bias. The margin of error might well be substantial, even very substantial.
69. With respect to each of the BCIF's ten constituent elements (indicators), no consideration is given as to how each indicator or the accompanying prompts has the potential to inform the judgement about whether the homicide was motivated by hate.
70. In fact, on this basis, there is good reason to doubt the reliability and validity of the BCIF as an

¹⁹ Ibid 3.

²⁰ Ibid 4.

²¹ Ibid 3.

indicator or measure of hate crime. When developing a scale predicting outcomes or selecting groups according to some criterion, the task is to identify a set of items that individually distinguish between the two groups/outcomes. This is ideally done empirically, here a mere inspection of the items offers sufficient cause to doubt their validity (validity determined by inspection is termed face validity.)

71. Consider the following components of the BCIF.
- (1) Differences': 'Victim is a member of a group which is outnumbered by members of another group ... ', 'Incident coincided with a holiday or date of significance ...':
 - (5) 'Previous existence of bias crime indicators ...', 'Victim was visiting a location ...';
 - (7) 'Motive of offender/s: 'victim perceived to be breaking from traditional conventions or working non-traditional employment';
 - (8) Location of incident: 'victim was in or near an area or place commonly associated with or frequented by members of a particular group', 'location of an incident has specific significance to the victim or POI group';
 - (9) 'Lack of motive': 'No clear economic or other motive for the incident exists';
 - (10) 'Level of violence': 'level of violence ... is greater than would be expected for a crime of that type'; 'weapons of opportunity are used ... ', 'The number of POI is greater than the number of victims and ... '.
72. So much is left open in these components to an investigator's personal interpretation as to whether gay hate is involved in the circumstances of a particular case. Are not many homicides not involving gay hate committed under these circumstances; and do not many victims of homicide not involving gay hate satisfy these criteria? It is not clear that each of these components is, in fact, a useful and valid identifier of hate crime. And if any of these are useful and valid identifiers of hate crime, it is also not clear that they have been stated appropriately so that they can be easily and consistently applied by an investigator using the BCIF.
73. Time may matter too. There appears to be no evidence in the report of the researchers considering how the indicia of hate crime may change over time. The period of 25 years – the time scale of this report – is a long time. Thus, for example, the character of armed robberies changed greatly between 1975 and 2000 due to prevention strategies adopted by potential targets and perhaps the greater role of drug addiction as a motive for offending. The nature of gay hate crimes may similarly have changed over the course of that time period. The criteria by

which gay hate crimes are identified may need to change over time as a result.

How was the BCIF used to identify a motive of 'gay hate'?

74. The Strike Force's use of the BCIF lacked sufficient specificity, clarity, and rigour. This necessarily impacted adversely on the level of reliability and validity. Furthermore, validity cannot be assessed independently when what was taken to be indicative of bias is so obscure. As discussed further below, this may affect:

- (1) the determination of bias in respect of each of the ten indicators in a particular case; and
- (2) the decision process for each case by which the combination of positive clues for each indicator and across the ten indicators is deemed to reveal bias, or not reveal bias (noting that the reasons for the decision taken in this second stage may not be explicit).

75. I say this for the following reasons. First, it is not clear how each indicator is specifically linked to the concept of bias. It starts with the Strike Force's definition of 'bias crime indicator': "objective facts, circumstances, or patterns attending a criminal act or acts, which, standing alone or in conjunction with other facts or circumstances, suggest that the offender's actions were motivated, in whole, or in part, by any form of bias."²² More specificity is required with respect to what precisely might suggest bias and how it might suggest bias. The 'indicators assist investigators systematically' (to determine) – but how do they do this?²³

76. As an example, the Coordinating Instructions state:

"For example, if the offender is recorded in police files as associating with persons known to have assaulted young gay men, then the investigator may mark Bias Crimes Indicator 4 (Organised Hate Group) as being relevant".²⁴

The language of '(M)ay mark' creates uncertainty. It begs the question of under what criminal circumstances of a crime a 'mark' would be made validly. The reader does not know at a sufficient level of detail the types of case fact that would be deemed indicative or not of bias. The same reasoning applies to the language of 'associating': questions arise such as the frequency of interactions, the nature of these interactions, the bigoted group's attraction to

²² *Strike Force Parrabell Coordinating Instructions* (n 16) 3.

²³ *Ibid.*

²⁴ *Ibid.*

the offender, and so on.

77. This problem is somewhat exacerbated by the Report's phrases relating to the role of hate in the crime. Thus, 'offender's actions were motivated, in whole or in part, by any form of bias' (above) is variously stated:
- (1) played a significant role;²⁵
 - (2) 'contributing factor';²⁶
 - (3) 'evidence indicative of';²⁷
 - (4) 'motivated by'; 'motivating factor';²⁸ and,
 - (5) anticipating the judgement of bias across the ten indicators, 'a bias indication does not have to establish that a bias ... was the main motivating factor'.²⁹
78. The qualification 'significant' in the first phrase as against the lack of any qualification in the second, third and fourth phrases, back to the potentially conflicting qualification in the fifth phrases, surely raises doubts as to meaning. This raises the question: 'How significant a role must bias play in order to amount to a hate crime?' This question is made more pertinent since bias did not have to be the principal motive for the case to be identified as a bias crime.³⁰
79. Perhaps this is no more than loose expression? However, consistency and rigour are to be preferred.
80. One instance of a vague operational definition alone may or may not be problematic in the context of a case. But it is a bedevilling problem for the BCIF's reliability (and validity) that definitional vagueness is to be found across multiple indicators, such as those identified above at [71].
81. There are additional related problems. What is the reader to make of 'by any form of bias' in the definition of a 'bias crime indicator', referred to above at [75]? This may be no more significant than a slip by the researchers in failing to adapt the US definition encompassing all

²⁵ Ibid 2.

²⁶ Ibid.

²⁷ Ibid 3.

²⁸ Ibid.

²⁹ Ibid 4.

³⁰ Ibid.

types of hate. However, this sort of looseness is always distracting and potentially misleading and distorting.

82. Finally, if an investigator does not know in general terms why a specified indicator may be indicative of gay hate, it surely makes it harder for the investigator to conclude whether that indicator is in fact indicative of gay hate in the circumstances of a particular case.

83. In light of these shortcomings, it is too open for reasonable minds to differ on whether gay hate was alive in the circumstances of a particular crime.

84. Now to the second part of the decision required for the classification of a case: the decision process for each case by which the combination of the individual positive clues across the ten indicators was deemed to reveal bias, or not represent bias. The fact is, the Report says little about the Strike Force's decision making in respect of this. It says no more than that,

“an overall conclusion will be made referring to each relevant indicator and the relevant evidence; comments will be recorded in order to clarify each finding; a bias indication does not have to establish that a bias was the only or main motivating factor behind an action contributing to their death”.³¹

85. With respect to this, the fundamental question for the investigators was said to be: “Is there [in this case] evidence of a bias crime?”³² A problem arises because of the large gap in content between what is covered by each of the BCIF's ten elements and the assignment of bias in regard to that aspect of the circumstances of the case (see preceding discussion). So, for example, in regard to the first indicator, ‘Differences’, (say) one difference deemed significant was identified. But we ask, on what basis does this represent bias? And as for each indicator, more so for the process of combining the evidence and bias attributions across the ten to arrive at an overall judgement. Questions again arise, for example:

- (1) Could the decision of bias on one element carry the decision?
- (2) Could less patent evidence of bias across several indicators serve to carry the day?
- (3) How might hate on several indicators combine to determine gay hate as a motive in the case? The process may be in effect largely additive – hate on two or more indicators

³¹ Ibid 4.

³² *Strike Force Parrabell Final Report* (n 7) 21.

sums to reach a threshold – or interactive – hate on two or more indicators is greater than the sum of the individual attributes.

86. More guidance is needed in these two aspects of the classification according to hate – each indicator and overall case. Too much was left to intuition. Nonetheless, let it be clear, there must always be room for intuition in the application of definitional guidance; a balance between specificity and generality. In general terms, erring in respect of the former will bring the error of exclusion; erring in respect of the latter invites the error of overinclusion. Where the balance best lies will depend on the complexity of the concept being measured and the nature of the data used for this purpose.

Differing standards of proof

87. In relation to the characteristics of reliability and validity, there is a further difficulty. Four intuitively sensible categories were created for the purpose of classifying cases involving hate, namely, ‘Bias Crime’, ‘Suspected Bias Crime’, ‘Not Bias Crime’ and ‘Insufficient Information’. For each case, an assessment in respect of each of these categories is made for each of the ten indicators separately and for the ten considered together. But the standard of proof differs across the two; somewhat puzzlingly, no reason being given.
88. The assessment of bias for each of the ten indicators was to be made according to the standard of ‘beyond reasonable doubt’.³³ This standard is not easy to interpret amid a messy evidentiary scene. It is a fine judgement, one learned by long experience in putting criminal circumstances before courts. The absence of this experience invites inconsistent judgements across independent raters, leading to poor reliability of the BCIF. Along with this, it sets a high evidentiary bar. While this minimises the risk of finding bias where there is no bias, it increases the risk of missing bias where it is actually present. This may set up the study to provide an underestimate of the incidence of bias among the 88 cases.
89. In comparison, the standard of ‘on the balance of probabilities’ is applied to the conclusion on bias as a motivating factor in a case considered in its totality.³⁴ This seems a more appropriate standard in view of the ‘messy’ character of the evidence. But is the damage already done in light of the ‘beyond reasonable doubt’ standard adopted for the individual indicators? Because of the opaqueness of the overall judgement of bias for each case (see preceding discussion),

³³ *Strike Force Parrabell Coordinating Instructions* (n 16) 4.

³⁴ *Ibid.*

the question must be left open.

90. A final observation on the standards of proof for the four bias categories. The standard in regard to the first, second and fourth are clear (notwithstanding their problematic application) and internally consistent, but the third, ‘not a bias crime’, is left open and stands apart. For this category there is no accompanying standard of proof to differentiate it from the former category.

Lack of information in relation to inconsistency

91. Evidence of the lack of clarity in the minds of the Strike Force team is apparent in the Report’s description of how the team went about making the required judgements. BCIFs were completed by individual investigative police officers, and subject to individual reviews both in batches as they were submitted and in one final sweep. For the former there were several teams and for the latter a final review team. The aim of these reviews was to ‘ensure consistency in methodology and conclusions’.³⁵ No information is given on the degree of inconsistency review teams were presented with at each stage of the review process. The review process may have been rigorous and elaborate but there is not sufficient information to confirm this. The review process would have been expected to improve rigour by increasing the coherence – internal consistency – of the classificatory process. However, the review process cannot make the BCIF as an instrument inherently more reliable.
92. In addition, the perceived need for these multiple reviews raises the possibility that the degree of *initial* inconsistency was not inconsiderable, and with this evidence of the unreliability of the BCIF as an instrument for measuring bias. This process of *review* itself did not ensure a high degree of consistency of the process. The actual level of unreliability could only be estimated by the process being *independently repeated* on the same cases and by different personnel, and then comparing the two sets of case bias assessments. This is what was required to determine whether the BCIF was sufficiently reliable as an instrument for the identification of bias.

Numerical analysis of data by Strike Force

93. The results of the analysis are presented across 15 tables.³⁶ Before proceeding to the most

³⁵ Ibid 4, 13, 14.

³⁶ *Strike Force Parrabell Final Report* (n 7) 23-37.

important observations on the presentation of the results, less significant matters bearing on clarity warrant mention. The points are made by way of questions:

- (1) Did being heterosexual as a victim or non-heterosexual as an offender exclude the possibility of the case being classed as involving hate bias?³⁷
- (2) What is to be made of the finding that in 21 of the cases the offender was found to be, or identified as, gay or bisexual?³⁸
- (3) Nine cases were classified as motivated by 'homophobia'³⁹ – one more than the number classified as 'bias crime'?⁴⁰ But is this not the fundamental question, 'Is there evidence of a bias crime?'⁴¹ What explains the discrepancy?
- (4) The finding of robbery being identified as the principal motive in 32 cases,⁴² presumably did not preclude the case being identified as a bias crime, since it was stated that bias did not have to be the principal motive for the case to be identified as a bias crime.⁴³ So what is to be made of this finding? In any case, the principal motive might have been robbery, but perhaps these particular people were targeted because they were thought of as 'fair game' because they were gay; if so, hate bias would have been an active factor warranting of itself the 'bias' label. Against this, the gay victims may have been targeted for robbery because they were seen as soft targets, not because of their sexuality, or alternatively targeted because it was thought that as gay they would not want to involve the police in light of supposed police prejudice.

94. These secondary queries about the results aside, now to the critical table, the table showing the numbers and percentages of cases falling in each of the four bias categories.⁴⁴ The numbers and percentages are Bias (8, 9%), Suspected bias (19, 22%), No bias (34, 40%), and Insufficient information (25, 29%).

95. To convey the potential extent of the problem of bias in these 86 cases, it would be a mistake

³⁷ Ibid 28-29.

³⁸ Ibid 29.

³⁹ Ibid 25.

⁴⁰ Ibid 37.

⁴¹ Ibid 21.

⁴² Ibid 37.

⁴³ *Strike Force Parrabell Coordinating Instructions* (n 16) 4.

⁴⁴ *Strike Force Parrabell Final Report* (n 7) 24.

to ground a summary conclusion on the figure of 9% for the Bias category. Why might it seriously distort the picture? Just consider the figures for the other three categories.

96. First, 'No bias'. This represents no evidence of bias on the basis of the Strike Force's understanding of hate bias both as a concept and, more importantly, as this notion was perceived by the team in the 88 cases. The latter in particular is somewhat obscure and may well not fit the understanding of hate bias in the gay community and/or the broader community. This figure based on other understandings might be higher; it might be significantly lower.
97. Second, with regard to the cases classified as 'Suspected bias' and 'Insufficient information', not insignificant numbers of these cases, particularly those in the former category, might have been deemed to fall in the 'Bias category' had there been additional information.
98. Where does that leave the reasonable reader? Accepting an estimate of gay-bias somewhere well above 9% but almost certainly below 100% of the 88 cases.
99. The results must be expressed as a range due to the degree of potential error and uncertainty associated with the Strike Force's methodology.
100. The importance of expressing the results as a range is underpinned too by this observation of the Strike Force team: "Unrest has continued within the LGBTIQ community [regarding] bias motivation related specifically to these 88 deaths".⁴⁵ This observation appears to refer to the period before the Parrabell Review, but nonetheless seems to be a similarly important consideration today.
101. It would be gross error, in light of the potential range of cases involving hate crimes (see above at [98]), to take the figure of 9% of cases involving gay-hate as representing the 'headline' finding of this research. That error may have potentially unfortunate personal and social consequences and may be distressing to members of the LGBTIQ community.

Conclusions

102. On the basis of my review of these matters, the following conclusions are drawn.
103. The term 'gay' is not clearly defined, leaving it somewhat open as to how what the Strike Force called 'gay' relates to the label 'LGBTIQ'. The Strike Force, in planning their research method,

⁴⁵ Ibid 18.

do not appear to have engaged the 'gay' community. If they in fact did consult the gay community, there is no evidence in their Report of this affecting their thinking or the indicia they adopted to identify gay hate in the circumstances of these cases. In this way, the Strike Force at once potentially ignored one aspect of their secondary objective and perhaps sacrificed validity in their measurement of gay hate (the principal objective).

104. The choice of the BCIF was not soundly based and cannot be taken to be adequate. It is not shown how it is appropriate for the research question. The indicators forming part of the BCIF were developed in relation to possible hate crimes generally; they were not developed for the purpose of identifying gay hate specifically. Along with this, there is no evidence that the BCIF was culturally applicable in the specific context of this study (the gay scene in Sydney, Australia).
105. The different standards of proof applicable to different parts of the BCIF process were problematic. In particular, the use of the standard "beyond reasonable doubt" for assessments with respect to each of the ten indicators unduly risked missing cases where bias was actually present.
106. The Strike Force adopted the BCIF without any evidence of its reliability and validity. In view of the team's process of reviewing cases for the purpose of consistency, the level of reliability of the BCIF may well have fallen below the level required of a quality behavioural measure. This was to be expected in light of the problems associated with Strike Force's methodology for identifying gay hate.
107. With respect to validity, an analysis of its face validity pointed to low validity. Moreover, low reliability necessarily undermines validity.
108. The preceding three problems render the BCIF a crude instrument for present purposes and the accuracy of any conclusion about the incidence of bias very uncertain.
109. The Strike Force's reporting of their analysis of the case data is too obscure. This applies to the use of the BCIF in identifying gay hate as a factor in these cases and, with this, the process of classifying the cases according to the presence of bias. Two consequences following from this are:
 - (1) the reader has no means of assessing the soundness (validity) of the team's judgements of bias, leaving the reader with no more than uncertain conclusions; and
 - (2) it fails a basic requirement of a social science research project, namely, the opportunity for independent researchers to replicate the actual study.

110. The pivotal role of the BCIF in this study represents faux science; it imparts a false sense of research rigour and validity.
111. The Strike Force's reporting of their analysis opens the way for the 'headline' conclusion to be: 'Only 8 of 88 cases positively found to involve 'gay' hate bias.'⁴⁶ This would represent a seriously misleading conclusion. More generally, having regard to the present review of the method, the Strike Force's Report provides no sound basis for any conclusion about the incidence of gay-hate bias in the 88 deaths. The 'true' figure may be quite low; the 'true' figure might be alarmingly high. We are none the wiser as a result of this research.

G. Question 2: What is your view as to the appropriateness of the different methodology used by the academic team, having regard to those objectives?

112. My assessment of the academic team's research methodology follows a logical sequence, in three parts. First, background matters are examined, including:
- (1) the definitions of hate and gay adopted by the academic team;
 - (2) the team's reasons for rejecting the BCIF for present purposes; and
 - (3) the team's alternative framework, a classificatory scheme intended to offer a more nuanced and valid interpretation of the circumstances of the 88 cases of suspected homicide on the 'list'.
113. Second, an assessment is made of the classificatory scheme as an adequate and sound instrument for identifying gay-hate crime, having regard to its actual use by the academic team.
114. Third, the academic team's numerical analysis of the incidence of gay-hate-motivated crime by way of this framework is evaluated.

Background Matters

Definitions

115. The academic researchers state that they "largely agree" with the police definition of bias crime.⁴⁷ Nonetheless, they do formulate their own definition of bias crime⁴⁸:

"(a) *expresses* a categorical animus (directed at a person or group on the basis of

⁴⁶ The Strike Force Report does not present their conclusion this way.

⁴⁷ *Strike Force Parrabell Final Report* (n 7) 81.

⁴⁸ *Ibid* 81-83.

his/her perceived identification with a *vulnerable* group).

(b) produces an act that *intentionally*, by way of criminal predation on the basis of that categorical animus, causes harm to that person or group.

(c) is mitigated or aggravated by an offender's contemporaneous associations that are linked by a commitment of *denunciatory non-identification* with the vulnerable person or group".⁴⁹

116. This definition excludes the qualification inserted by Police i.e. '... motivated, *in whole or in part* ...'. In practice this may constitute a very significant difference. Since this is part of the one discussion, it would appear to be a conscious decision on the academic researchers' part. This suggests that to the Police a crime in which hate was a secondary motive was still deemed a hate crime. The question arises whether the academic team required more in this respect; if they indeed did – the reference to 'financial' identified as the principal motive suggests this may be the case – they would find fewer hate crimes: see below at [129(4)]. In regard to the academic team's elaboration, hate also requires some degree of overt expression of animus. The question must be asked, 'Why?' Again, if this informs the team's interpretation of evidence, they will find fewer incidences of hate crime. The academics offer a second elaboration; it is in respect of 'association', which is more appropriately dealt with in the following discussion of their classificatory framework.
117. As to their understanding of 'gay', they say little, other than it involves sexuality and gender, presumably meaning sexual orientation and the public flouting of gender norms. In respect of the LGBTIQ label, there is no reason not to take it as all-embracing; note, their apparent inclusion of transgender women and men who cross-dress as part of a discussion of nuancing sexuality and gender.⁵⁰
118. It is problematic that the academic team appears to eschew victim perceptions of gay hate so readily.⁵¹ In my view, these perceptions informing what victims understand to constitute gay hate must be attended to in any understanding of the phenomenon. Their observations should not be accepted uncritically in respect of this, since this may result in the over identification of gay hate. Rather, this input will be approached with the attitude that the gay community may

⁴⁹ Ibid 82-83.

⁵⁰ *Strike Force Parrabell Final Report* (n 7) 108.

⁵¹ Ibid 78.

be aware of manifestations of gay hate not in the awareness of individuals who are not in this community. The academics' argument is that gay victims are in no better position than anyone else to determine gay bias in the circumstances of a crime.⁵² Certainly, the observations of other witnesses to a crime are potentially valid, but the gay community's experiences and thoughts must be central to the understanding of what constitutes a gay-hate crime. It should not be forgotten that many in the gay community will be victims and associates of victims of anti-gay violence.

Rejection of BCIF and adoption of alternative classificatory framework

119. The academic team does not regard the BCIF as adequate to the task for two reasons. First, it is accompanied by no evidence regarding its reliability and validity; this includes the constituent elements themselves and the assessments made in relation to these elements considered individually and together. I agree with the academic team's assessment in this regard: see above at [63]-[68].
120. Second, the academic team asserts that as an instrument for the identification of hate, the BCIF is not soundly based or sufficiently nuanced. Assessments not based on key elements of bias as a measure are too crude. With this, they say, the BCIF does not differentiate between different expressions of hate, which reveal variation in motive and have different implications for certainty of classification and for social policy and, consequently (from this perspective), the potential to 'over-categorise' gay bias.⁵³
121. To address these problems, the academic team makes three categorical distinctions. They are: Level of predation (Proactive/Reactive); Offender's hate association (Associative/Non-associative); Victim's identity (Anti-gay non-paedophile/Anti-gay paedophile).⁵⁴ The third is relevant to the definition and identification of gay-hate; the first two are not, rather a distraction. All three, each in their own way, may incline the reader to underestimate the incidence of gay-hate or less than fully appreciate the extent of gay-hate the 88 cases truly represent. The problem arises from the reasoning in both being grounded on social policy.
122. First, consider the 'predation' and 'association' distinctions. The academics state that an offender who offends in some way in association with like-minded gay-haters and who has

⁵² Ibid 82.

⁵³ Ibid 71, 73.

⁵⁴ Ibid 88.

actively sought out his victim (cf. an offender absent these associations who has reacted violently to some sort of perceived homosexual provocation) can be regarded as more dangerous,⁵⁵ and thus presumably warrants greater attention for the purposes of enforcement and prevention. Moreover, the label gay-hate can be assigned to their crime with greater certainty.⁵⁶ Whether an offender is more or less dangerous is surely irrelevant to the determination of whether or not the crime involves gay-hate. There is a further complication. 'Association' is deemed a factor not only when the offender offends when in communication with like-minded people but, alternatively, where an offender associates with like-minded people.⁵⁷ Thus, a 'reactive' crime could warrant the label 'association', yet the academics' classificatory scheme does not include/allow a 'reactive-association' classification.⁵⁸ Indeed, one might suppose that a person who associates with like-minded people (i.e. those motivated by gay hate) may be predisposed to react negatively to a gay sexual advance. That could be classified as 'reactive-association', if the academic team had adopted such a classification. Nevertheless, is not such a person dangerous, especially having regard to his 'associations'? The problem arises because motive and circumstances are not sufficiently differentiated.

123. And regarding the academics' observations on certainty; the certainty of a crime only becomes significant for identification of hate bias if the classificatory scheme includes a 'Suspected' category, which the academics' does not.
124. Now to the distinction that the academics make in respect of anti-gay bias and anti-paedophile animus.⁵⁹ The argument here is that it is important to distinguish between a pure anti-gay hate and an anti-paedophile hate. The argument is that these two categories sometimes get conflated in the minds of perpetrators, as a result producing a particularly potent hatred. The academic team insist that the two must be distinguished on public policy grounds, namely, that not making the distinction is to perpetuate the slander that they are one and the same people. This is surely problematic. If the aim is to identify gay-hate crime, then motive counts whether or not it is based on a misperception. And in situations where a paedophile is involved, the 'gay' component of the bias may not be expressed either at or outside of the scene of the crime. To

⁵⁵ *Strike Force Parrabell Final Report* (n 7) 83.

⁵⁶ *Ibid* 84.

⁵⁷ *Ibid* 89-90.

⁵⁸ *Ibid* 90.

⁵⁹ *Ibid* 84-85.

the offender this may be self-evident.

Assessment of the alternative classificatory framework, as applied

125. The purpose of the conceptual analysis by the academic team, in the preceding section, is to present and explain their understanding of the problem, and to identify the factors critical to this understanding. It also allows the reader to evaluate the soundness of this understanding and is the basis of any assessment of validity. The definitions of the factors identified are also critical in this analysis. These definitions should be used to determine how to measure or assess the relevant phenomenon, here, to identify the presence of gay hate in the circumstances of the 85 cases. However, the academic team do not take this step; they offer nothing more specific, no operational definitions. The team broadly favours the use of '(o)bjective facts, circumstances, or patterns attending a criminal act...' as bias crime indicators,⁶⁰ yet offers nothing more than examples of classified cases, described in no more than several lines.⁶¹ Wide is the gap indeed between the behaviour described in the definitions and the detailed facts arising in the specific circumstances of the individual case.
126. Thus, when the researchers first individually assessed the cases according to their classificatory scheme they, as individuals, had inadequate guidance as to how the hate was manifested according to the scheme in the circumstances of actual cases.⁶² An elaboration on the academic team's thinking process used to classify cases is offered in Appendix C. I found it unenlightening; it lacked (but certainly required) an illustration of its application in a particular case.⁶³ Problematic reliability among members of the research team was inevitable; so it proved to be. Though no numerical measure was made, we can discern this from their account of the coding process.⁶⁴ Thus, cases were independently coded and then group reviewed in an effort to reach consensus. Moreover, the subsequent independent coding on the revised instrument also required a subsequent concordance consultation. The academic team's use of a review process would have been expected to improve rigour by increasing the coherence of the assessments. However, the group review process cannot make the application of the alternative framework

⁶⁰ Ibid 81.

⁶¹ Ibid 89, 95-96.

⁶² Ibid 90.

⁶³ Nonetheless, it matters little, since it relates to their classificatory framework which is, in my opinion, irrelevant to the determination of the presence of the motive of gay hate in a case.

⁶⁴ Ibid 90.

inherently more reliable.

127. Again (see above at [86]), let it be clear, there must always be room for intuition in the application of definitional guidance; a balance between specificity and generality. In my view, the academic team's definition erred in the former respect in that they left too much room for intuition.
128. The practical problems arising for the researchers in the complexity of the case circumstances are well illustrated in the discussion of this matter in their Report.⁶⁵ Thus, the researchers state: '... there was often difficulty applying it [the carefully (sic) defined instrument] to complex case details'.⁶⁶ Here the researchers demonstrated the problems of identifying hate and classifying cases according to the categories in their framework ('proactive' and 'associative'), and in respect of cases of anti-gay bias and anti-paedophile bias; also in respect of anti-gay bias where the class of crime (robbery) invites an alternative interpretation as the nature of the hate involved (financial).
129. Text extracted from this discussion illustrates the issue.
- (1) Associative: 'In some cases a person privy to a crime was either not directly involved in the crime or predominantly a witness to an event that they did not or could not anticipate. Consistent with the meaning of "association" in the literature and our model, we coded these as "Non-associative". To class a crime as "Associative" we wanted to see evidence of two or more people conspiring on a shared prejudice to cause harm'.⁶⁷ This reasoning seems to fit the first part of their definition of "association" – when an offender offends when in communication with like-minded people – but not the second part – when the offender associates with like-minded people: see above at [122].
 - (2) 'Proactive': '... difficulty of coding 'reactive' or 'proactive' (as though they are pure and mutually exclusive categories) ...'.⁶⁸
 - (3) 'Anti-paedophile bias': 'Confronted with a variety of animuses at play, the academic team were persuaded to classify this crime as Anti-paedophile bias. ... This

⁶⁵ Ibid 100-104.

⁶⁶ Ibid 100.

⁶⁷ Ibid 101.

⁶⁸ Ibid.

categorisation may appear slightly flawed in its logic, but the academic team wanted to capture the element of paedophile hatred captured in this most complex case.⁶⁹

- (4) Robbery: The Report does not consider that a robbery may be explained by anti-gay hate rendering the victim as a 'worthy/not unworthy' target for robbery.⁷⁰
130. As noted above at [129(3)], the academic team observed: "... a variety of animuses at play ... the team wanted to capture the element of paedophile hatred ...". This leads one to infer that where both anti-gay and anti-paedophile hatred were at play, there was a prevailing team disposition to preference the classification of paedophile over gay.
131. It appears that any offenders who took it as self-evident that most paedophiles were in fact gay (an assumption unexpressed at the scene of the crime) would not be counted as anti-gay. The academic team's discussion of the case of Tuckey illustrates the problem "... when a crime is categorised (in this case as Anti-paedophile because the killer used the terms 'poofter' and 'rock spider' to describe the victim)".⁷¹ The academic team reach this conclusion, although in the offender's mind the two were apparently closely associated and the animus expressed at the scene was joint, and although the academic team acknowledges that at the time of the offending the presumption was certainly common among the community of potential perpetrators.
132. In these circumstances one might have thought the more reasonable approach would have been to categorise all such cases as anti-gay, unless there was clear evidence to the contrary. Of course, this would carry the error of over-inclusion, just as their approach carries the error of under-inclusion. Yet in the milieu of the two animuses being not uncommonly conflated, the former error might be expected to be the smaller of the two. It would appear that the academics' approach allows for a case where minor anti-gay bias would be trumped by anti-paedophile bias in the categorisation of cases.
133. The academic team proceeded on the assumption that a case falling in the 'Associative' category was more likely to involve hate.⁷² Thus, in all this, the question arises as to how their considerations about certainty actually influenced their decisions in the allocation of cases to

⁶⁹ Ibid 104.

⁷⁰ Ibid 102-103.

⁷¹ *Strike Force Parrabell Final Report* (n 7) 105.

⁷² Ibid 70-71.

the four bias categories – how it might tip the balance between the multiple considerations.

134. In view of this complexity and in the absence of well-considered operational definitions, there is only so far a researcher can go in producing an acceptable level of inter-rater consistency (reliability). Nonetheless, their instrument (framework) for categorising cases cannot be regarded as ‘carefully defined’, as they suppose;⁷³ it did not go far enough in identifying the indicia of hate in the circumstances of actual cases.
135. There is no independent means of assessing the degree of dissensus present, and hence the magnitude of the problem for reliability. But one might reasonably suppose it was not insignificant: see above at [126]. If there was dissensus among a group who had already ‘worked’ the data prior to the coding process, an independent group would have great difficulty trying to replicate the study. There are problems here too for validity. Without knowledge of what is meant by hate in the circumstances of these cases, how can outsiders know what the study actually found, assess its soundness, or apply the results to public policy?

Numerical analysis of data by Academic team

136. The results of the analysis are presented across three tables/figures (2,4 and 6).⁷⁴ The critical table/figure is Figure 2 (the ‘associative-non associative and ‘proactive-reactive’ distinctions in Figures/Tables 4 and 6 being regarded as not relevant to the identification of gay hate).⁷⁵ Table 2 shows the numbers of cases falling in each of the four categories, namely, ‘Anti-gay bias’, ‘Anti-paedophile animus’, ‘Insufficient information’ (this includes cases for which the researchers suspected bias), and ‘No evidence of bias’.⁷⁶ The numbers (and percentages) for these categories are, respectively, 17 (**20%**), 12 (**14%**), 33 (**39%**), and 23 (**27%**). Upon combining the first category (Anti-gay bias) and second category (Anti-paedophile animus) to form one Anti-gay bias category (see above at [124]) the figures for Anti-gay bias, Insufficient information, and No evidence of bias, become, respectively, 29 (**35%**), 33 (**39%**), and 23 (**27%**).
137. To convey the potential extent of the problem of bias in these 85 cases, it would be a mistake to ground a summary conclusion on the figures of 20% or 35% for the ‘bias’ category. Why might it seriously distort the picture? Just consider the figures for the other two categories.

⁷³ Ibid 100.

⁷⁴ Ibid 92, 94, 95.

⁷⁵ Ibid 70-72.

⁷⁶ Ibid 92.

138. First, 'No evidence of bias'. This represents no evidence of bias on the basis of the academic team's understanding of hate bias both as a concept and, more importantly, as this notion was perceived by the academic team in the 85 cases. The latter, in particular, is somewhat obscure and may well not fit the community's understanding of hate bias or, importantly, have given due regard the gay community's understanding of hate bias. This figure based on other understandings might be significantly higher; it might be significantly lower.
139. Second, with regard to the cases classified as 'Insufficient information', a not-insignificant number of these cases, particularly those where bias was suspected, might have been deemed to fall in the 'Bias' category had there been additional information.
140. Where does that leave the reasonable reader? Let us as a starting point accept the figures for 'bias' and 'no bias'. Then we can do no better than reaching an estimate of gay-bias somewhere above 20-35 % but almost certainly below 100 % for the 85 cases. And movement would be expected between the three categories, in accordance with other understandings of gay hate.
141. The soundness of the conclusions thought of as a range is also underpinned by the degree of error associated with the academic team's methodology.
142. The importance of the range is underpinned too by this observation of the Strike Force team: "Unrest has continued within the LGBTIQ community [regarding] bias motivation related specifically to these 88 deaths".⁷⁷ (This observation appears to refer to the period before the Parrabell Review, but nonetheless seems a similarly important consideration today.)
143. As with the conclusion reached by the Parrabell officers (see above at [101]), it would be gross error, in light of the potential range of cases involving hate crimes to take the figure of 20% of cases involving gay-hate (the figure for the 'Anti-gay bias' category) to represent the 'headline' finding of this research. This would, again, have potentially unfortunate personal and social consequences, and may be distressing to members of the LGBTIQ community.

Conclusions

144. On the basis of this review of these matters, the following conclusions are drawn.
- (1) The term 'gay' is not clearly defined, leaving it somewhat open as to how what the academic team call 'gay' relates to the label 'LGBTIQ'.

⁷⁷ Ibid 18.

- (2) The academic researchers do not use an actual instrument for identifying gay hate but have a broad definition of hate. They do develop a framework for identifying the various circumstances in which gay hate is expressed and as it is related to motive, but this adds little to identifying the incidence of gay hate in the cases before them.
- (3) The academic team, in planning their research method, do not appear to have engaged the 'gay' community, in this way perhaps sacrificing validity in their measurement of gay hate.
- (4) The academic team do not provide evidence supporting the reliability of their assessments of gay hate in the cases. The acknowledged need for review suggests that reliability of the initial application of the framework to detect hate was not reliable.
- (5) The academic team's reporting of their assessment and identification of gay hate in the case data is too obscure. They do engage a definition of hate, but the gap between this definition and the manifestation of hate in the circumstances of the individual cases is too great. Two consequences follow from this. They are:
- i. the reader has no means of assessing the soundness (validity) of the team's judgements of bias, leaving the reader with no more than uncertain conclusions, and
 - ii. it fails a basic requirement of a social science research project, namely, the opportunity for independent researchers to replicate the actual study.
- (6) The academic team's reporting of their tabular analysis opens the way for the 'headline' conclusion to be: 'Only 17 of 85 cases positively found to involve 'gay' hate bias.'⁷⁸ This would represent a seriously misleading conclusion.
- (7) More generally, having regard to the present review of the method set out above, the Academic Report provides no sound basis for any conclusion about the incidence of 'gay'-hate bias in the 85 deaths. The 'true' figure may be quite low; the 'true' figure might be alarmingly high. We are none the wiser as a result of this research.

⁷⁸ The Academic Report does not present their conclusions this way. Nonetheless, one of the academics' reports of their research – an article in an academic journal – surely risks conveying this message: de Lint and Dalton, 'Anatomy of Moral Panic: The "List of 88" and Runaway Constructionism', (2021) 29 *Critical Criminology* 723. Thus, at 724: "Both academic investigators and the police task force determined that most of the homicides on 'the list' either did not involve gay-bias or that there was insufficient information to make a determination".

H. Further Commentary on a Comparison of the Parrabell and Academic Reports

145. Little more can be concluded than that the number of the cases on the 'list' in which gay hate played a not-insignificant role, but not necessarily a principal role, will range from a relatively small but not-insignificant minority to a quite, perhaps even very, substantial majority. This is broadly true of both the Police findings and the academic findings. There can be nothing more definite by way of conclusion.
146. First, while at first blush we might be tempted to conclude from the percentage of cases in each study found to involve hate, that no more than a relatively small minority of the homicides involved gay hate, this would be patently unsound. One problem arises from there being too many cases for which it was not possible to make a determination of their gay-hate status. If there had been more information available about these cases, the incidence of gay hate may have been found by both teams to be much higher; then, perhaps, one team may have found much hate, another little hate.
147. Second, there are the problems of reliability and validity. For both the Police and academic teams, it is tempting to conclude that the final classification of cases represented a somewhat Procrustean uniformity. Various members of each team had their individual views arising from their own takes on the facts of each case. What underlay the final composite view of each team, let alone the individual views, we know not. And how these understandings differed within and between each team we know not. Thus, it is not open to the reader to evaluate whether they accept the understanding of hate behind each team's classifications. Indeed, each final determination may represent no more than an averaged somewhat compromised view.
148. Third, even if there had been a rough agreement between the two studies on the number of cases involving bias and the number not involving bias, this would have represented no more than a weak finding of agreement. The Report includes no data showing whether the Police and academic teams allocated the same cases to the same categories.
149. Fourth, any high level of agreement between the two studies would come with uncertainty. The question would have to be asked, 'Does this represent true agreement – fully independent agreement?'
150. For the Police, the academic research was to be independent.⁷⁹ Yet the Academic Report refers

⁷⁹ Ibid 14.

to extensive conversations with the Police team, which inadvertently may have inclined them to the Police understanding of the expression of gay hate in the 88 cases. What is to be made of the following passage?

“... academic team engaged in carefully measured debates [with the police] about each individual case in the interests of being thorough, consistent and precise ... [this] allowed the academics to develop a more nuanced understanding of the logic that underpinned the categorisation decisions of the Strike Force”.⁸⁰

151. This is not to say that the Police thinking necessarily infected the academic team’s thinking at all, let alone significantly, but the question must be raised.
152. Nonetheless, in one respect Police understandings did constrain the academic team’s capacity to identify gay hate in the 88 crimes. This arises from the quality of the raw data with which the academic team had to work. These data were the ‘case summaries’ (apparently the BCIF forms completed by the police investigators). It must be presumed that any prejudices or limitations of understanding of what constitutes gay hate would infect these case summaries and exclude or distort information which the academics might have considered material to the finding of gay hate.

Signature:



Name: Associate Professor Austin Lovegrove

Date:

27-1-23

⁸⁰ Ibid 57.