MUCH ADO ABOUT CRIME
Much ado about crime
Much ado about crime
Chapters on Psychology and Law

Miet Vanderhallen
Geert Vervaeke
Peter Jan Van Koppen
Johan Goethals
Foreword
Foreword

Psychology and law, as an applied scientific discipline, was rather slumbering until the end of the sixties. Since then, the interest in the domain of psychology and law became prosperous, first in the United States and somewhat later in Europe. After several national conferences, the first European Conference on Psychology and Law was held in the Netherlands in 1988, which was the cradle for the establishment of the European Association of Psychology and Law (EAPL) at Nuremberg in 1990 (formally founded in 1992 in Oxford). One of the aims of the EAPL is to organize yearly conferences in which the state of the art in research and practice is presented.

At the 10th conference in 1999 in Dublin, The Irish President, Mary McAleese, posed three challenges for the development of psychology and law: Internationalism, application and interdisciplinarity. The 12th European Conference on Psychology and Law in Leuven, Belgium, September 14-17, 2002 pursued to contribute especially to the last mentioned challenge, the so-called interdisciplinary approach. This was realized through keynote lectures that were commented on by either colleagues from other disciplines (such as sociologists, criminologists, psychiatrists, lawyers and practitioners working in the field) or other psychologists.

Along with the traditional topics within the domain of psychology and law, special attention was paid to fear of crime, dangerousness, and violence. This focus originated from the society’s growing interest towards these topics. Besides that, fear of crime, violence and dangerousness are actually at the top of the political agenda in many European countries. Since research on psychology and law can – without any doubt – contribute to this debate, it would have been a lost opportunity if these topics were lacking.

During the conference, the Dean of the Faculty of Law, Prof. dr. Frans Vanistendael, added another challenge for the domain of psychology and law. In screening the programme, he noticed that most of the research is directed to penal law. In his view, many other areas of law could profit from empirical research based on the methodological and theoretical know-how within psychology and law.

The 12th European Conference on Psychology and Law attracted many participants from Europe and beyond. From the many interesting contributions, a selection from a range of papers was made. This resulted in a selection of chapters, which give an overview of the diversity in the conference program. In the introductory chapter the magistrate’s role is described from the perspective of one who is also a practicing forensic psychologist. It also reviews research on relevant issues including: safeguarding witnesses at trial, situational and witness factors, motor vehicles, video evidence, novel sentencing options and magistrate’s decision making. The other chapters are organised in three parts: fear of crime, witnesses, and victims and offenders. The first part deals with topics as the confusion with the conceptualisation and operationalisation of fear of crime, measurement error, and preventive actions to decrease perceptions related to the fear of crime. The second part focuses on the reliability and accuracy of (eye)witnesses. In the third part studies are described with either a focus on offenders or a focus on victims.

We are indebted to many people for the organisation of the conference. First of all we would like to thank the members of the scientific and organising committee for their contribution before and during the conference. Furthermore, the success of the conference was a result of the hard work of the following collaborators of the Faculty of Law of the Catholic University of Leuven: Ann Claes, Barbara Cloet, Anne Groenen, Stefaan Pleysier, Marjoly Sintobin, Katrien Vandecruys, Marc Van De Plas. We would also like to thank Marc Van de Sijpe, Kevin Van de Sijpe, François Opdebeeck, Maria Janssens and Nadine Jansens for their generous help during the conference. At last of course, we would like to thank all the authors for their constructive collaboration.

The title of this volume, Much Ado About Crime, refers to Shakespeare’s tale Much Ado About Nothing (see the Lambs' tales from Shakespeare). Shakespeare’s Much Ado About Nothing deals with love, deceit, counterfeiting, impression management, grief, shame, detection of truth, and remorse. Furthermore the word “nothing” in the title was pronounced as “noting” in Shakespeare’s time referring to the actions of observing, listening, writing and noting of many players in the tale. Both the themes and actions in this tale are of major importance.

in the applied research domain of psychology and law. But *Much Ado About Crime* can also be understood in two different ways. First, crime has become one of the most important political topics in society. No one will deny that citizens, the media as well as politicians are preoccupied with the topic of crime. Secondly, it is clear that there is a lot to do about crime, also in the domain of research and practice in psychology and law. The chapters in this volume do demonstrate how much indeed is done by legal psychologists in the study of crime.

Miet Vanderhallen  
Geert Vervaeke  
Peter J. Van Koppen  
Johan Goethals  
Leuven, January 2003
Contents
Foreword

Introductory chapter
A foot in both camps: on being a forensic psychologist and a magistrate
Davies, G.M.

Part 1  Fear of Crime

Chapter 1
Fear of crime, personal safety and well-being: a common frame of reference
Bilsky, W.

Chapter 2
An international perspective of community-based factors
Carrington, J. & Pfeifer, J.

Chapter 3
Assessing measurement invariance in ‘fear of crime’ instruments
Pleysier, S., Vervaeke, G., Goethals, J. & Billiet, J.

Chapter 4
Fear of crime in transitional Poland
Czapska, J. & Wojcikiewicz, J.

Part 2  Witnesses

Chapter 5
Laypeople’s criteria for the discrimination of reliable from non-reliable eyewitnesses
Arce, R., Fariña, F. & Seijo, D.
Chapter 6 The influence of field dependency on eyewitness accuracy in free and cued recall
*Emmet, D., Clifford, B.R. & Gwyer, P.*

Chapter 7 “I’ll Never Forget the Sinking Ferry” How Social Influence Makes False Memories Surface
*Granhaug, P.A., Strömwall, L. A. & Billings, J.F.*

Chapter 8 Confidence-accuracy calibration for absolute and relative judgements
*Weber, N. & Brewer, N.*

Chapter 9 The verbal overshadowing effect: in search of a chimera.
*Clifford, B.R.*

Part 3 Victims and Offenders

Chapter 10 Family violence and animal abuse by juveniles The intragenerational cycle of violence
*Baldry, A.C.*

Chapter 11 Rape responsibility attributions as a function of victim attractiveness and race
*Jalbert, N.L. & Sniffin, H.*

Chapter 12 Telling and detecting true lies Investigating and detecting the lies of murderers and thieves during police interviews
*Vrij, A. & Mann, S.*

Chapter 13 Violence risk assessment An anchored narrative approach
*Hart, S.D.*

Chapter 14 Violent and non-violent young offenders A comparison of demographics, multi-problem profiles, and psychopathic traits
*Corrado, R.M., Cohen, I.M. & Marino, F.*

Chapter 15 The role of animal cruelty in the prediction of dangerousness
*Medway, C. & Howitt, D.*
Chapter 16  Criminal careers of juvenile psychopaths  
*Spidel, A., Greaves, C., Hervé, H. F., Yuille, J. C. & Hare, R.D.*

Chapter 17  Psychopathy and deceptive motivations in young offenders  
*Spidel, A., Hervé H.F., Greaves, C., Cooper, B. S. & Hare R.D.*

Chapter 18  The spousal assault risk assessment (SARA) guide and the psychopathy checklist-revised (PCL-R)  
*Thomas, L.A.*

Chapter 19  Cognitive bias and judicial decisions  
*Fariña, F., Arce, R. & Novo, M.*

List of contributors
LIST OF CONTRIBUTORS

ARCE, R., Department of General and Social Psychology, University of Santiago de Compostela, Spain.

BALDRY, A. C., Department of Social and Developmental Psychology, Faculty of Psychology, University of Rome ‘La Sapienza’, Rome, Italy. National Institute of Statistics, Italy.

BILLIET, J., Department of Sociology, Catholic University of Leuven, Belgium.

BILLINGS, J. F., University of Texas, El Paso, USA.

BILSKY, W., Department of Psychology, University of Münster, Germany.

BREWER, N., School of Psychology, Flinders University, Adelaide, Australia.

CARRINGTON, J., Canadian Institute for Peace, Justice, and Security. Department of Psychology, University of Regina, Saskatchewan, Canada.


COHEN, I. M., Simon Fraser University, Vancouver, Canada.

COOPER, B. S., University of British Columbia, Vancouver, British Columbia, Canada.

CORRADO, R., Simon Fraser University, Vancouver, Canada.

CZAPSKA, J., Department of Sociology of Law, Jagiellonian University, Cracow, Poland.

DAVIES, G. M., School of Psychology, Leicester University, UK.

Victims and offenders

FARÍÑA, F., General Psychology, University of Vigo, Spain.

GOETHALS, G., Department of Criminal Law & Criminology, Catholic University of Leuven, Belgium.

GREAVES, C., University of British Columbia, Vancouver, British Columbia, Canada.

GRANHAG, P. A., Göteborg University, Sweden.

GWYER, P., King Alfred College, Winchester, UK.

HARE, R. D., University of British Columbia, Vancouver, British Columbia, Canada.

HART, S. D., Simon Fraser University, Vancouver, Canada.

HERVÉ, H. F., University of British Columbia, Vancouver, British Columbia, Canada.

HOWITT, D., Department of Social Sciences, Loughborough University, Loughborough, Leicestershire, UK.

JALBERT, N. L. Department of Psychology, Western Connecticut State University, Danbury, CT, USA.

MANN, S., Department of Psychology, University of Portsmouth, Portsmouth, UK.

MARINO, F., Kwantlen University College, Canada.

MEDWAY, C., Department of Social Sciences, Loughborough University, Loughborough, Leicestershire, UK.

NOVO, M., Department of Social Psychology, University of Granada, Spain.

PFEIFER, J., Canadian Institute for Peace, Justice, and Security. Department of Psychology, University of Regina, Saskatchewan, Canada.

PLEYSIER, S., Department of Criminal Law & Criminology, Catholic University of Leuven, Belgium.

SEIJO, D., Department of Developmental Psychology, University of Granada, Spain.

SNIFFIN, H., Department of Psychology, Western Connecticut State University, Danbury, CT, USA.

SPIDEL, A., University of British Columbia, Vancouver, British Columbia, Canada.


VANDERHALLEN, M., Department of Criminal Law & Criminology, Catholic University of Leuven, Belgium.

VAN KOPPEN, P. J., Faculty of Law, University of Antwerpen, Antwerpen, Belgium.

VERVAEKE, G., Department of Criminal Law & Criminology, Catholic University of Leuven, Belgium.

VRIJ, A., Department of Psychology, University of Portsmouth, Portsmouth, UK.

WEBER, N., School of Psychology, Flinders University, Adelaide, Australia.

WOJCIECHOWICZ, J., Department of Criminalistics, Jagiellonian University & Institute of Forensic Research, Cracow, Poland.

YUILLE, J. C., University of British Columbia, Vancouver, British Columbia, Canada.
INTRODUCTORY CHAPTER
A Foot in Both Camps: On being a Forensic Psychologist and a Magistrate

WOLFGANG BILSKY

‘Magistrate’ is a term widely used in Europe to refer to those who preside over lower level criminal courts. However, magistrates in England and Wales, have an unusual and distinctive role and background. Edward III created the first ‘Justices of the Peace’ in 1361, who were required to ‘pursue, arrest, take and chastise offenders’ in their locality. This alternative title for magistrates is still in use today, though their role and function has evolved over the centuries. Magistrates are no longer involved in policing, but have acquired many other responsibilities, such as the regulation of gambling and public houses and the approval of search warrants by the police (Gibson, Gordon & Wesson, 2001; Grove, 2002). Unpaid and part time, selected for their fair-mindedness rather than their legal knowledge, lay magistrates preside over 96% of all the criminal cases heard in the English courts every year. Along with the citizen jury, they represent the still radical precept in English common law that, as far as possible, all those accused of crimes should be tried and judged by their peers. In this paper, I shall describe the role and function of magistrates in England and Wales and summarise some of the research that has been conducted into their procedures and penalties. I will also comment on such issues from the perspective of a forensic psychologist who also serves a magistrate.

ROLE AND FUNCTION

How does one become a magistrate? There is inevitably, a form to fill in: you can nominate yourself or be nominated by others. There are at least rounds of interviews conducted by the local Advisory Committee who are charged with ensuring not only that persons appointed have sufficient integrity and respect for the law
Introductory chapter

to sit in judgement on others, but also that, as far as possible, the ‘bench’ of magistrates (the term for the group of magistrates attached to a particular court) reflects the profile of the community as a whole. Surveys of the backgrounds of the 30,000 odd magistrates in England and Wales suggest that such representation is achieved on some, but not all indicators. Just over half of all magistrates are women and strenuous efforts are made to recruit from the ethnic minorities by advertising in the ethnic press and campaigns aimed at local communities; in 1999, some 8.8% of all magistrates in training were from the ethnic minorities (Gove, 2002). Political persuasion is also a relevant factor and all applicants must declare the political allegiances to ensure a balance of opinion. While magistrates can be appointed from the age of 30, most are middle-aged, with retired persons figuring disproportionately (all magistrates must retire from the bench at 70). Likewise, in terms of social class, while there are certainly many magistrates who would describe themselves as working class, the middle classes are over-represented. The reasons for this may lie with the demands of the position: in addition to sitting a minimum of 26 days a year in court, there are training days and regular bench meetings. Many private-sector employers are reluctant to countenance such absences and ambitious young employees in any sector will calculate that evidence of such divided loyalties is unlikely to win favour or promotion.

All magistrates on appointment must swear on oath: “to do right to all manner of people... without fear or favour, affection or ill will”. After some preliminary training, they begin sitting on cases within months of appointment. Magistrates hear cases in groups of three—a chairman and two ‘wingers’. Chairs must have at least five years experience in addition to special training. A justices’ clerk, a legally qualified professional who sits within whispering distance of the magistrates, advises on matters of law. There are no juries in Magistrates’ Courts and decisions will be made on the day, often after a brief adjournment to enable them to confer: the aim is always cheap and swift justice for crimes which carry a maximum penalty of up to six months’ imprisonment. If trials are included, some 97% of all cases end in a guilty verdict: the great majority of defendants plead guilty or are found guilty after a short trial, typically lasting a day or less. District Judges, professionals who sit on their own and can dispense with the need for legal advice, supplement magistrates in some busy inner-city courts.

Magistrates try a wide range of petty crime: what are termed ‘summary matters’. The staples of most Magistrates’ Courts are motoring offences: all but the most serious will be dealt with by fines and penalty points. Rather than the process known as ‘totting’—if the 12-point barrier is breached within a three-year period, then a driving ban will be imposed. However, even serious crimes, ‘indictable offences’ such as murder, rape, conspiracy and firearms offences, which are always tried before a judge in the higher or Crown Court, begin with an appearance by the accused in the Magistrates’ Court. For example, Ian Hunter, one of the defendants in the recent Soham murder case, appeared before magistrates in Peterborough to be committed for Crown Court trial. In addition, there are a range of so-called ‘Either way’ offences, which can be tried either before a judge
and jury in the Crown Court or before magistrates. Currently, if magistrates agree that their sentencing powers are sufficient and the defendant agrees to have his or her case heard in the lower court, then Justices of the Peace will try such cases. Offences falling into this category include the worst burglaries, criminal damage, supplying drugs and serious assaults.

In addition to the normal court hearings, there are also specialised courts presided over by experienced magistrates who receive further specialised training. Youth Courts, for defendants aged between 10 (the age of criminal responsibility in England and Wales) and 18 years of age, attempt to prevent the development of criminal careers through early intervention and treatment. The YOT (youth Offender Treatment) teams made up of a social worker, probation officer, police officer and health or education representatives reflect this philosophy. While prevention may be the main role of the Youth Courts, deterrence is not forgotten and indeed, the sentencing powers of these courts are actually greater than for the adult courts. Family Courts deal with a range of family matters, including the custody and adoption of children and the fallout from family break-up, which can include the issuing of non-molestation and supervision orders. The emphasis is upon conciliation whenever possible and cases generally take much longer than the swift justice of the mainstream Magistrates’ Court.

Monitory penalties form by far the most common penalty imposed by magistrates, with some 76% of all convictions resulting in fines. Community-based penalties are available for defendants who commit more serious crimes that merit a penalty that falls short of imprisonment. Some 9% of cases ended with defendants undertaking periods of either community punishment (garden maintenance for the elderly; refurbishment of community buildings and facilities) or community rehabilitation (the old probation order) or on occasion, both. Despite their existing powers of sentence, magistrates send just 4% of all convicted defendants to prison. In a further 1% of cases, magistrates dealing with ‘either way’ offences send the defendant to the Crown Court for sentence, because they believed their own powers are inadequate given the magnitude of the crime, an eventuality that all defendants are warned of, prior to agreeing to have their cases heard in the lower court. Recently, changes have been mooted by the Auld Report (2001) in relation to ‘either way’ offences that would have removed the right of defendants to determine their mode of trial, a change strongly resisted by civil liberties groups. A likely compromise will see magistrates having their powers extended to deal with ‘stronger’ cases: perhaps those carrying a sentence of up to one year’s imprisonment.

DAILY LIFE ON THE BENCH

My own bench covers a predominantly rural area, including England’s smallest county of Rutland. Traffic offences certainly loom large in our business, with one of Britain’s busiest trunk roads-the A1- running through our area. Other com-
Introductory chapter

mon types of offences are casual violence and criminal damage, frequently associated with drink or drugs. For many defendants, ‘Saturday night fever’ seems to have an inevitable consequence of Tuesday morning court appearance. For those on their first offence, a small fine or even agreeing to be ‘bound over to keep the peace’ (a caution available to magistrates since the 14th Century) may suffice. More serious offences, particularly those resulting in assaults on police officers or hospital staff, may result in the ordering of a pre-sentence report from the probation service, foreshadowing a community sentence or even imprisonment.

In addition to such staples, the court printout is often leavened by some exotica. Recently, there have been a number of stalking cases, a topic that crosses over into my academic interests (Sheridan & Davies, 2001). The 1997 Protection from Harassment Act gave magistrates powers to issue Restraining Orders to inhibit defendants from contact with the victim with the option of a fine of up to £5,000 and/or six months imprisonment. My experience is very much in line with the first national survey reported by the Home Office (Harris, 2000). This reported that, contrary to media portrayals, just 2% of all cases involved stalking by strangers; the rest involved prior acquaintances of which partners or ex-partners represented a significant group (41%). Unwanted telephone calls, sometimes ‘silent’ calls where the caller replaces the receiver as soon as the call is answered, form one of the most common forms of stalking behaviour (Budd & Mattinson, 2000), though some stalkers resort to more ingenious ways of harassing their victims. In one case, the stalker contacted the local police officer who dealt with domestic violence, reported that his victim was being stalked and urged her to go around late at night, just to check that she was alright!

Living in a country area, our court presides over a significant number of cases involving allegations of cruelty to animals, many brought independent of the Crown Prosecution Service by animal charities such as the Royal Society for Protection of Cruelty to Animals (the RSPCA). Such ‘private prosecutions’ can on occasion throw up difficult issues. In one case, the RSPCA produced a video of an apparently dead sheep, lying in a pile of farmyard detritus including the bodies of other sheep, but whose limbs were palpably moving. The former was on trial for neglect of his animals, a charge that can lead to severe fines or imprisonment in animal-conscious England. Both parties called veterinary surgeons as expert witnesses to support their case, the prosecution arguing that the animal had been prematurely left to die, while the defence argued that the animal was suffering from hypothermia. They pointed to the very low temperatures on the day in question, that the animal could have appeared dead in the field when collected by the farmer, but the inherent warmth of the midden then thawed the animal out sufficiently for some semblance of life to return. After some deliberation, we found the farmer not guilty.

In making our decision in that case, we were greatly aided by the presence of an experienced farmer on the bench. The sheer diversity of background of magistrates (no more than 15% can come from any one professional group) ensures
that there is a wealth of experience available in relevant cases. What of my own expertise as a forensic psychologist in issues around evidence? Some of my fellow magistrates thought I should be able to tell who was guilty by scrutinising their non-verbal behaviour alone! They were clearly not familiar with the literature on lie detection, which demonstrates the low diagnostic value of such cues (see Vrij, 2000 for a review). As a very junior magistrate, I am only too aware that I have a great deal to learn, but occasionally, knowledge of research findings is useful, for instance in urging colleagues not to put too much weight on confidence and eye contact in assessing testimony from the witness box. More useful that any research knowledge perhaps is the skills that come from a long experience of reading and assimilating information from witness statements and briefs, a form of expertise which lawyers and solicitors possess to an advanced degree, but which has yet to be studied by psychologists in any systematic way.

RESEARCH ON MAGISTRATES’ PROCEDURES AND PENALTIES

As an academic I am aware that there are large numbers of studies on eyewitness testimony and identification, a cottage industry to which I have contributed more than my fair share. However, there is relatively little research on how witness evidence is received in the courts or how judges evaluate such evidence. As regard research on the effectiveness of Magistrates’ Courts in England and Wales, there are many evaluation studies dealing with the impact of various sentencing initiatives and procedural innovations, many only available in government reports of limited circulation. Such research is largely atheoretical, in keeping with the evidence-based approach adopted toward law-enforcement in the United Kingdom (Davies, 2002). There are a few experimental studies, designed to isolate and identify factors thought to be critical to the decision-making processes of the Magistrates’ Courts. In Baddeley’s terms (1979), there is some cognitive applied research (evaluation studies, using methodology and methods of analysis pioneered in laboratory research), but little applied cognitive research (experimental studies informed by wider issues derived from theory). A selection of these evaluation and experimental studies are reviewed below.

SAFEGUARDING WITNESSES AT TRIAL

In the adversarial system of justice, most crucial evidence for the prosecution and the defence is heard live in open court on the day of the trial in keeping with the traditional emphasis on oral evidence. Thus, it is important that all witnesses ‘come up to proof’: what they have to say in evidence mirrors what they have reported earlier to the police or to their lawyer and that their evidence is put across in a clear and lucid manner. The collapse of trials through the failure of witnesses to give coherent and convincing evidence from the box has been a concern for successive Governments. One area of concern has been the intimidation of prosecution witnesses by defendants, their family or friends. Tarling, Dowds
and Budd (2000) report the results of a major survey of witness experience, which found that 8% of all witnesses reported being overtly threatened if they gave evidence. The source of the threat was the accused in 84% of the incidents reported, most frequently in association with a domestic incident. Magistrate's existing powers enable them to attach special Bail conditions to an accused to inhibit or prevent interference with witnesses: an accused can be ordered not to communicate with a given witness and not to go with a fixed distance of his house or place of work, under pain of being committed to prison prior to trial. Such powers are perhaps less effective when intimidation is conducted by family or friends of the accused. Additional assistance in dealing with such problems is contained in new Guidance, written by a team of psychologists and lawyers (Home Office, 2002).

The 1999 Youth Justice and Criminal Evidence Act introduced a number of new special measures to assist judges and magistrates to minimise the possibility of intimidation at court. They include the provision of screens and access to CCTV (known as a 'Live link') to enable the witness to testify and be cross-examined from a remote location (typically another room within the court complex) out of sight of the accused in court. With agreement of the Court, the 1999 Act also permits the use of pre-recorded testimony; experiments are to be conducted into the feasibility of pre-recorded cross-examination, which could obviate the need for vulnerable witness to attend court at all. Crown Courts in England and Wales pioneered the use of CCTV in cases involving children as witnesses, where the use of Live links is now almost universal (Davies, Wilson, Mitchell & Milsom, 1995). Provision of CCTV in Magistrates’ Courts for the use of vulnerable or intimidated witnesses of all ages is now underway. Such links can be used not only for vulnerable or intimidated witnesses of all ages, but also for routine hearings prior to trial, where the accused is required to appear at court to make an application for bail or to be committed for trial. Until now, prisoners and their escorts have had to be physically transported to courtrooms, often at some distance from their home prison, to appear in court for a five-minute hearing, an expensive and time-consuming procedure. Under new arrangements, such hearings can be conducted via a video circuit linking the prisoners’ goal to the courtroom. An evaluation of the effectiveness of these new procedures found widespread acceptance by prisoners: some 73% preferred the new arrangements to the traditional personal appearance (Plotnikoff and Woolfson, 1999). There is some interesting psychological research to be done on the impact of personal versus video appearances of prisoners on the perceptions of magistrates’ and their readiness to grant bail under the old and new procedures.

**SITUATIONAL AND WITNESS FACTORS**

The influence of situational and witness factors on the accuracy and completeness of testimony is one of the oldest and best-studied areas of forensic psychology. I do not propose to reprise those findings yet again (see Wright & Davies,
2000 for a recent review). However, from the magistrates’ perspective, there are issues that bear heavily on the credibility of witnesses, which are commented upon by lawyers in court, but on which forensic psychology has yet to explore with any thoroughness. The first and most obvious is the impact of drink and drugs, singly or in combination on witness testimony. It is not always appreciated that not only the accused, but also many witnesses may be intoxicated at the time of crimes typically heard in the Magistrates’ Courts: affrays; vandalism; common assault and domestic violence to name but a few. In one case, a defendant’s legal representative sought to excuse his clients’ evasive replies by claiming he had drunk alcohol after ingesting a ‘beta-blocker’ drug. Our current research knowledge is such that neither the lawyer nor the court could say with any authority whether this would have an adverse effect on his testimony and if so, how much of an effect (the so-called calibration problem: Wells & Turtle, 1987). Some experimental studies on the impact of alcohol on eyewitness testimony demonstrate that it is possible to study such issues in an ethically acceptable, but forensically realistic way. This sparse literature also demonstrates that any simple dose-response relationship is modulated by other factors such as levels of stress and arousal (Read, Yuille & Tollestrup, 1992). More studies exploring the impact of so-called recreational drugs on witness performance are also urgently required (Yuille, Tollestrup, Marxsen, Porter & Herve, 1998).

MOTOR VEHICLES AND THEIR SPEED

Witnesses frequently claim from the box that they saw the accused driving at \( x \) miles an hour. How accurate are such estimates likely to be? Should the estimate of a police patrol officer carry more weight than a member of the public? And can such estimates be influenced by extraneous factors, such as the characteristics of the car and driver and the consequences of a resultant accident? One classic experimental study on peoples’ ability to estimate vehicle speed suggests that accuracy can be very poor in some circumstances and how questions are asked can significantly influence not merely the estimated speed, but also the witness’s perceptions of the severity of a subsequent accident (Loftus & Palmer, 1974). Some recent research from my own laboratory has involved members of the public reading brief accident reports and driver statements of the kind frequently heard in the courts. Large and reliable effects were found on estimates of speed, position on the road and culpability for the accident, depending upon the stated age, type and colour of the cars involved (Davies and Patel, 2001). It would be good to think that magistrates’ training and experience would make them immune to such stereotypical thinking, but this remains to be demonstrated rather than assumed.

VIDEO EVIDENCE IN THE COURTROOM

Figures suggest that in the United Kingdom as a whole there are now some 1.5 million CCTV cameras in operation of which over half a million are exclusively
Introductory chapter

concerned with security surveillance (Parliamentary Office of Science and Technology, 2002). Such cameras scan city streets, the interiors of buses, department store exits, cash points, and all other areas where public security is a source of concern (Norris & Armstrong, 1999). Feed from such cameras is routinely recorded, and frequently is, submitted as part of the evidence against an accused, particularly in public order offences. Such evidence can be enormously useful in determining the guilt or otherwise of an accused: high quality recordings under good lighting conditions can provide an objective record, which facilitates the often difficult task of balancing one account against another. One example concerned an alleged assault on a female police officer by a young man outside a nightclub. When heard in the Magistrates’ Court, no video record was produced and the magistrates, having heard the conflicting accounts of both parties, found the accused guilty, but of a minor assault. The case then went to a District judge for sentence. In the meantime, the police had located a video record of the entire incident which, when played at court, substantiated the original prosecution case and totally discredited the defence. However, the judge’s hands were tied by the decision of the earlier court and he was unable to give the accused the sentence that his behaviour might well have deserved.

Quite apart from its role in truth determination, there is a need for further research on the impact of video evidence on magistrates’ sentencing behaviour. Does the graphic nature of video evidence lead to heavier sentences than those accorded on the basis of written or oral statements? One piece of research suggests this may well be the case. Groups of magistrates were shown either video recordings of a series of actual incidents or verbal statements covering the same incidents. Magistrates who saw the video recordings awarded higher tariff sentences on average compared to those who were confined to purely verbal evidence (Chenery, Henshaw, Parton & Pease, 1988). Yes the sheer impact of good quality video should not blind us to the dangers of an over-reliance on such evidence. Poor quality recordings may well raise serious issues as to the identity of those shown in the recording. Experiments conducted in my own laboratory have shown that even when an unfamiliar suspect is seen on video in good lighting conditions, volunteers correctly identified her from a photographic array on only 29% of occasions. A further 65% of volunteers selected an innocent person on a (Davies & Thason, 2001). Clearly, the Turnbull guidance, which applies in Magistrates’ Courts as well as Crown Courts, deserves to be taken seriously where questions of mistaken identity are raised.

NEW SENTENCING OPTIONS

A fourth area of research concerns assessments of the impact and effectiveness of the sentencing procedures now available to magistrates. Recent years have witnessed a proliferation of such Orders, usually stimulated by popular public concern, however ephemeral. For example, there are now special orders for restraining football hooligans, dangerous dogs, anti-social neighbours and sex offen-
A FOOT IN BOTH CAMPS: ON BEING A FORENSIC PSYCHOLOGIST AND A MAGISTRATE

...ders living in the community. Most involve community-based penalties, designed to resolve issues early to prevent more serious offences occurring. Two major innovations in recent years are electronic monitoring or ‘tagging’ and Drug Treatment and Testing Orders, the impact of both of which have been subjected to extensive research.

The aim of electronic monitoring is to restrict a convicted offender’s movements in a controlled and systematic way. Typically the offender will be confined to their homes for fixed periods through the day or night in an effort to disrupt offending patterns, such as robbery, driving or drug offences, which take place regularly within specified hours. It can also be used as part of an early release scheme for prisoners to ease their integration back into the community. In both cases, it enables offenders to be monitored regularly by the probation service and to allow more intensive intervention than might otherwise be possible (Mortimer, Pereira & Walter, 1999). Convicted offenders who agree to be tagged are fitted with an electronic device that transmits a signal to a monitoring unit connected in turn to the domestic phone line. Signals are transmitted regularly and any move by the offender away from his home during the specified hours will be detected and subsequently investigated as a potential curfew violation. Potentially, tagging offers many of the advantages of regular surveillance and constraint associated with prison, without the cost and the disruption to the offender’s family life. Mortimer (2001) examined the use of tagging as part of an early release scheme and found no difference in rates of re-offending after 6 months for tagged prisoners, compared to those who completed their sentence in the normal way: both groups showed rates of re-offending of around 30%. These are encouraging findings, though it should be noted that those who agreed to be tagged were not a random sample, but had been subject to a risk assessment analysis. Other research has compared the likelihood of re-offending of a group sentenced to tagging orders compared to others serving other community-based penalties. Again, no differences were observed in re-conviction rates between the two samples after two years: the rate was a depressing 73% overall (Mortimer et al., 1999; Sugg, Moore & Howard, 2001). Thus the main advantage uncovered so far for this innovation appears to be one of reducing costs, rather any rehabilitative effect, though one study does suggest greater compliance with community-based rehabilitation programmes among those who agreed to a tagging order (Sugg et al., 2001).

Drug treatment and Testing Orders (DTTOs) are targeted at habitual drug users who are also offenders and aims to wean them off drugs through intensive monitoring and regular treatment. Once again, compliance by offenders is required in return for a reduced custodial sentence. Given the impact drug use has on overall rates of crime, the results have been encouraging. Turnbull, McSweeney and Hough (2000) report that the average expenditure on drugs by offenders on the Orders fell from £400 per week at the time of arrest to £25 per week, with a particular impact on the use of ‘hard’ drugs (crack/cocaine; amphetamines) as opposed to opiates. Reported rates of offending also fell dramatically over the survey period, though difficulties in monitoring those who left the programme before...
completion (anything from 28 to 60% in the areas studied) may have led to an overestimate of the effectiveness of DTTOs. Such orders are popular with magistrates, but their use is limited by costs: intensive treatment involves the services of specialist professionals and regular testing—often as high as 3 times a week.

MAGISTRATES’ DECISION-MAKING

Magistrates’ training emphasises the importance of structured decision-making in an effort to produce reliable and consistent verdicts. Chairmen are enjoined to consult with wingers and at trial, to report their decisions in a systematic written form, emphasising facts not in dispute, facts in dispute and the reasons for preferring one version of events over another. Sentencing is constrained by statute, with different crimes attracting a different range of tariffs, dependant upon the severity of the offence. Nonetheless, within the proscribed range, there is considerable scope for variations in precise sentence. Again, guidelines, produced by the Magistrates Association are available and widely used by magistrates in sentencing. These offer an ‘entry point’ for a particular offence and suggest aggravating and mitigating circumstances which could serve to increase or decrease the severity of an offence. Thus, offenders convicted of ‘driving while disqualified’ could end up in prison if they had an existing disqualification and had deliberately evaded detection. However, those offenders who established a real emergency, say a partner who had to be rushed to hospital in the absence of other transport, might face a fine or even a discharge. Having established the seriousness of the offence, magistrates should also consider any personal mitigation by the offender, such as displays of genuine remorse and evidence of cooperation with the police.

Despite the intensive training and written guidelines, large-scale regional variations in sentencing are still common in England and Wales. For instance, a fine for failure to possess a current licence to receive television programmes is one of the most common and low-level offences dealt with magistrates. It is straightforward offence—either you have a licence or you don’t—and might be expected to elicit consistent treatment in different courtrooms. However, figures recently released by Fines Enforcement Agency (Johnson, 2002) show major regional variations. At one extreme, offenders convicted in Wiltshire and Derbyshire faced average penalties of £199 and £179 respectively, while those committing the same offence in Leicestershire and Cumbria received fines of just £92. Magistrates clearly have some way to go in their quest for consistent sentencing. Some variation is acceptable and expected: fines in more prosperous parts of the country would be expected be higher than those levied in more deprived districts. Likewise, it could be argued that some offences should attract stiffer sentences in certain areas, where a particular type of crime is a cause of local concern: racially motivated attacks in inner-city areas might be one example. However, such special factors do not explain the variation that has been observed for virtually every type of offence dealt with by magistrates.
Can psychology offer any explanation for this variance in convictions and sentencing? Research on the impact of extraneous factors on sentencing by magistrates is at an early stage, but the results show promise. The appearance of defendants and the way they dress and present themselves in court may be a significant factor (Hedderman, 1990). Likewise, information inadvertently acquired as to the defendant’s previous record (Lloyd-Bostock, 2000) may also be another potentially distorting factor. Changes in the law can also help to improve consistency. Under the new Human Rights legislation, magistrates are required to state the reasons in open-court for their decisions in contested cases. This brings England and Wales closer in judicial practice to some European judiciaries, where this procedure is already well established. Our colleagues in the Netherlands have produced a powerful critique of this process in action and proposed a theory of ‘anchored narratives’, which they argue could serve to increase the consistency of decision making in the courts (Wagenaar, van Koppen & Crombag, 1993). Research based around the concept of anchored narratives might usefully be applied to the English Magistrates’ Courts.

CONCLUSION

As this paper illustrates, lay magistrates dispense a great deal of justice in England and Wales, but their roles and responsibilities are not widely known and research on their role and impact is not well publicised. With society’s newfound emphasis upon audit and accountability, the training regime for lay magistrates is now more intense, in an effort to produce a more professional and consistent approach to dispensing justice. The danger of such professionalism is that the time and commitment required will limit those volunteers who wish to become magistrates to the affluent and the retired and the courts will become less widely representative of the community; currently, magistrates may not mirror society but their class, educational and gender mix are infinitely more representative than say, Crown Court Judges (Dyer, 2002). As has been noted, procedures in Magistrates’ Courts have recently been amended in the light of Human Rights legislation, which has had the effect of offering some convergence in judicial procedure between the adversarial tradition of English common law and the inquisitorial approach of mainland Europe. Such convergence offers for the first time the opportunity for a pan-European research agenda for law and psychology to understand the mysteries of the courtroom.

REFERENCES


PART 1
Fear of crime
Fear of crime, personal safety and well-being: a common frame of reference

WOLFGANG BILSKY

I | INTRODUCTION

Interest in fear of crime has been manifold in the past. Politicians, representatives of the media as well as scientists from different disciplines like criminology, political sciences, or sociology have been concerned with this topic. In spite of the multitude of empirical studies conducted during the past decades, research in this domain has not been without problems (for a critical review of the literature, see Greve, 1998; Hale, 1996). Thus, Sessar (1992) stated with reference to social upheaval and to criminological research: “The terrific speed of development in Germany (and in Eastern Europe) produced the splendid saying: Words have already become obsolete when leaving our lips. The same may happen to criminological research, if ... the nature of the relation between radical change and criminality as a process were neglected” (p.138; translation by the author). This reservation applies still today, both to criminological research in general, and to research on fear of crime in particular. Additionally, two further caveats have to be added: namely, the definition of the variables under investigation in an adequate and unambiguous way and the interpretation of the data in the context of scientific theory. These caveats are in the focus of the present paper.

Psychologists, too, have become involved in research on fear of crime - often, if not mostly, because of their methodological know-how. However, there are good reasons for them to deal with this research domain from a theoretical perspective as well: Psychologists can rely on decades of research on fear and anxiety, on stress and coping, or on critical life events - to name but a few fields that may be reasonably linked to the subject matter under consideration. In the following, I will approach fear of crime from a slightly different, however related, perspecti-
To demonstrate this, I start from a broad, empirically based taxonomy of well-being. Next, some basic ideas of Facet Theory (FT) are sketched out, as applied by Levy and Guttman (1989) to the analysis of well-being, coping and related topics. Finally, the FT-approach to well-being is adapted to the analysis of personal safety and fear of crime. Data from a national survey conducted by the Criminological Research Institute of Lower Saxony (KFN) are used for illustrative purposes (Bilsky, Pfeiffer & Wetzels, 1993; Wetzels, Greve, Mecklenburg, Bilsky & Pfeiffer, 1995; Bilsky, 1996; Greve, 1998).

2 PERSONAL SAFETY AND WELL-BEING

At least at first glance, the fact that studies on fear of crime abound seems to underscore their relevance. From media research we know, however, that agenda setting has considerable impact on the salience of a particular problem like crime. The discussion of such a topic, whether in the public or in science, often develops a momentum of its own and can only partly be explained by the actual relevance of the underlying problem. Furthermore, social change or political transformations like those in Germany today are likely to give rise to (existential) fear and anxiety, concerning housing, unemployment, health care, etc. - at least for some part of the population. Such problems may undoubtedly challenge personal safety and eclipse other problems like (fear of) crime. It seems reasonable and necessary, therefore, to check whether the attention devoted to crime in the public debate and in science matches the importance attributed to it by the individual, that is, by the possible victim. What is needed, then, is a common frame of reference, which makes it possible to evaluate the supposed impact of crime relative to other problems from the individual’s perspective.

Following this line of reasoning, crime can be conceived of as one out of a multitude of potential stressors that threaten personal safety. These stressors are supposed to cause strain in the individual because of the perceived discrepancy between the desired and the real state of personal safety. In addition to this discrepancy, the amount of strain is also dependent on whether and to what extent the individual feels competent to cope efficiently with the respective situation.

From psychological research we know that the absence of strain and the availability of coping resources are central and defining features of subjective well-being.
Mayring (1991), for instance, contends that at least four components of well-being should be differentiated. As shown in figure 1, lack of strain is one of them. This interpretation is in line with victimological reasoning. Bayley (1991) states that people are victims if and only if “they have suffered a loss or some significant decrease in well-being unfairly or undeservedly and in such a manner that they were helpless to prevent the loss ...” (p. 53; italics by the author).

Figure 1: Aspects of well-being (cf. Mayring, 1991, p. 53).

These considerations form the basis for sketching out a framework within which to localize and evaluate (fear of) crime. Starting from some general information about FT, a conceptual system is presented which shows similarities and differences among well-being, coping and related concepts.

3 | A FACET APPROACH TO WELL-BEING

Facet theory (FT) is a (meta-) theoretical approach developed by Louis Guttman, which can be characterized by three features: Facet design, a companion set of multivariate statistical procedures and hypotheses of correspondence between design and data analysis (Borg, 1993; Borg & Shye, 1995; Shye, Elizur & Hoffman, 1994). Theory, as understood in this context, means “an hypothesis of a correspondence between a definitional system for a universe of observations and an aspect of the empirical structure of those observations, together with a rationale for such an hypothesis” (Levy & Guttman, 1989, p. 469).

Facet design enables the researcher to define his research interests formally in such a way that systematic data collection and data analysis are facilitated. In this context, a facet may be any way of categorising observations as long as the elements of the respective category are mutually exclusive. However, identifying facets supposed to be relevant with regard to analysing the current research pro-
Fear of Crime

...blem, is but a first step. As a second step, the relations among the facets as well as among their elements must be specified. This is accomplished by means of a *mapping sentence*. Such a sentence can be read from top to bottom like a sentence in ordinary language by combining the appropriate elements (1...n) of the different facets (A...Z) in order to identify a special case of the phenomenon under study. Whether or not the facets chosen are ‘useful’ for answering a particular research question has to be answered by data.

*Multivariate statistical procedures* usually applied in FT data analysis comprise a set of ‘soft’ methods that work “with few restrictions and assumptions, so that the substantive findings do not become loaded with mathematical structures and constraints that are irrelevant to the studied domain” (Borg & Shye, 1995, p. xi). Nevertheless, traditional (e.g., linear) procedures for analysing empirical data may be used as well. In the present context, ordinal Multidimensional Scaling (MDS) is of particular interest. MDS represents similarities, e.g., correlations between items, in low-dimensional space in such a way that higher similarities of any two variables correspond to smaller distances of their representations in space.

*Hypotheses of correspondence*, finally, refer to the relation between the definitional system (i.e., the facets and their mutual relation) and the empirical observations (data) as represented in multidimensional space. In analysing these relations, facets play specific roles (Levy, 1985): Depending on the type of contingencies supposed to exist between psychological variables, different partitions of data, i.e., regions, are expected to emerge in space representing these contingencies (e.g., wedge-like partitions, ordered bands, concentric regions, or combinations of these splits). It should be noted that regional hypotheses, as used in FT, usually refer to space that, in principle, has data points everywhere. This is so because they do not refer to a sample of concrete items but to the *universe*, i.e. the population of items under consideration (Borg & Shye, 1995; Levy, 1985).

Levy and Guttman (1989) used the FT-approach in order to bring two concepts together within one common framework - *well-being* and *coping*. According to these authors, well-being items assess satisfaction concerning the situation or the treatment of an individual or a social group in some life areas. Coping items, on the other hand, relate to (cognitive, affective, or instrumental) behaviour directed against a possible negative state of a life area of an individual or a social group. In terms of their definitional system, expressing satisfaction and coping can be regarded as two varieties of *adjustive behaviour* towards a situation. To provide a common framework for well-being and coping, Levy and Guttman proposed two major facets for defining the domain of adjustive behaviour items, the *mode* and the *directive* of behaviour. Figure 2 shows the respective mapping sentence for observations on adjustive behaviour.
By combining the elements of the two major facets A and B, different forms of adjustive behaviour can be distinguished. The following forms are of particular interest in the present context (Levy & Guttman, 1989; Borg, 1993): satisfied to be, i.e., well-being \{a_1, b_1\}, uneasiness to accommodate, i.e., strain \{a_2, b_4\}, and ability to try to attack, i.e., coping \{a_4, b_3\}.

To test the validity of their theoretical approach, Levy and Guttman specified the correspondence hypothesized between the elements of these facets and the regions of the MDS space that represent the correlations between adjustment-items. In other words, they formulated regional hypotheses that link the classes of their definitional system to regions of a representation space for the data (Borg, 1993, p. 117). This was accomplished by referring to the roles, facets usually play in partitioning multidimensional space (see Levy, 1985, for a comprehensive overview).

Without going into methodological detail (cf. Levy & Guttman, 1989; Borg, 1993), findings from ordinal MDS of survey data corroborate the hypothesized data structure: Areas of life (Facet E) plaid a polarizing role, cutting space into wedge-like regions that emanate from a common origin; primacy of environment (Facet D) plaid a modulating role, partitioning space in concentric bands around a common origin. Both facets appeared on the same two-dimensional projection in space, so that the resulting structure looked like a dartboard, i.e., a radex. Finally, the items of the directive facet (Facet B) came out ordered on an axis perpendicular to this radex. Figure 3 is a schematic representation of the partitioning of the MDS space according to these three facets of adjustive behaviour. This partitioning resembles more a cone than a cylinder. This is due to the fact that there is more spread at the well-being level than at the coping level. Or, in terms of the

---

<table>
<thead>
<tr>
<th>A: mode</th>
<th>B: directive</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_1 satisfied)</td>
<td>(b_1 be)</td>
</tr>
<tr>
<td>(a_2 uneasy)</td>
<td>(b_2 continue)</td>
</tr>
<tr>
<td>(a_3 motivated)</td>
<td>(b_3 attack)</td>
</tr>
<tr>
<td>(a_4 able to try)</td>
<td>(b_4 accommodate)</td>
</tr>
<tr>
<td>(a_5 likely)</td>
<td>(b_5 protect)</td>
</tr>
</tbody>
</table>

---

**Figure 2:** Mapping sentence for observations on adjustive behaviour
data, “well-being items have, on the average, lower intercorrelations than ‘attack’ (coping) items, with ‘continue’ items in between” (Borg, 1993, p. 126).

Figure 1: Schematic representation of the partitionings of the MDS space corresponding to three facets of adjustive behaviour (cf. Levy & Guttman, 1989, p. 471).

At least at first glance, notations in terms of Facet Theory look somehow cryptic and require getting used to. Nevertheless, FT proves extremely useful with respect to both, designing research in a transparent way, and analysing data efficiently. This is demonstrated next by applying FT to the analysis of personal safety and strain.

4 PERSONAL SAFETY AND STRAIN

Findings from preparatory studies conducted as part of our own research on fear of crime (Bilsky & Wetzels, 1994) support the view that the absence of stimuli, supposed to be a challenge to a balanced state of well-being, is characteristic of the lay concept of personal safety. Furthermore, incidents mentioned to threaten personal safety were related to life areas like those distinguished in the previously cited studies on well-being (Levy & Guttman, 1989; Levy, 1990).

Because of the similarity of connotations of well-being and of personal safety, we expected that empirical analyses that relate to the latter concept would reflect this similarity, too. Stated differently, strain caused by threatening stimuli is supposed to be but a special form of adjustive behaviour, namely uneasiness to accommo-
date \{a_2, b_4\} in terms of the above mapping sentence (figure 2). Therefore, asking people to what extent their feelings of personal safety are threatened by different stressful events, and submitting their answers to multidimensional scaling, was assumed to yield structures similar to those known from research on well-being.

As we were especially interested in problematic conditions of life, the respective element of facet C in the mapping sentence of Levy and Guttman was expanded. According to Young (1991, p.30), three primary injuries can be identified as causing major distress to victims: financial injury or loss, physical injury or loss, and emotional trauma. Correspondingly, injury was included as a separate facet in our own mapping sentence, comprising material, physical, and psychological injuries as its elements.

Two other facets were directly adopted from Levy and Guttman’s work, the environment and the life areas facet. As regards environment, primary (social) environment is usually defined by close and intimate face-to-face interaction with other social agents, e.g., family members, peer groups, etc., as well as personal (emotional) involvement. Secondary environment, in contrast, is mostly used as a complementary (rest) category, referring to less frequent or less direct social contacts and experiences. The life areas facet distinguishes between different categories of concern, like health, work, economy, etc. Figure 4 displays the modified mapping sentence of adjustive behaviour, which focuses on strain as a defining characteristic of subjective well-being.

![Figure 4: Mapping sentence of strain and personal safety.](image-url)
Fear of Crime

The usefulness of this approach for localising crime relative to other stressors of personal safety was tested in two studies, a pilot study (N=213) and a national survey (N=11,116). Stressors were taken from literature on victimization and from our preparatory studies. Each of them was classified \textit{a priori} to empirical analysis according to the three facets of our mapping sentence, as shown in table 1. The pilot study and the national survey differed only insofar, as the crime-related stressor remained unspecified in the first (‘crime’) as opposed to the second study (‘assault’; ‘theft, robbery, fraud’).

Table 1: Stressors of personal safety, investigated in a pilot study (N=213) and a German national survey (N=11,116).

<table>
<thead>
<tr>
<th>Item (abbreviation)</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-criminal stressors</strong></td>
<td></td>
</tr>
<tr>
<td>loss of job</td>
<td>a1 b2 c2</td>
</tr>
<tr>
<td>inflation and economic crisis</td>
<td>a1 b2 c3</td>
</tr>
<tr>
<td>uncertainty of life annuity (pension)</td>
<td>a1 b2 c3</td>
</tr>
<tr>
<td>severe illness</td>
<td>a2 b1 c1</td>
</tr>
<tr>
<td>war</td>
<td>a2 b2 c1</td>
</tr>
<tr>
<td>natural disaster</td>
<td>a2 b2 c1</td>
</tr>
<tr>
<td>crisis in the health service</td>
<td>a2 b2 c1</td>
</tr>
<tr>
<td>separation of children</td>
<td>a3 b1 c4</td>
</tr>
<tr>
<td>environmental damage</td>
<td>a2 b2 c1</td>
</tr>
<tr>
<td>to become dependent on others</td>
<td>a3 b1 c4</td>
</tr>
<tr>
<td>accident</td>
<td>a2 b1 c1</td>
</tr>
<tr>
<td>loss of apartment</td>
<td>a1 b1 c5</td>
</tr>
<tr>
<td>chance</td>
<td>a1 b3 c6</td>
</tr>
<tr>
<td>family fights and anger</td>
<td>a3 b1 c1</td>
</tr>
<tr>
<td><strong>Criminal stressors</strong></td>
<td></td>
</tr>
<tr>
<td>crime (pilot study)</td>
<td>a4 b3 c6</td>
</tr>
<tr>
<td>assault (national survey)</td>
<td>a4 b3 c6</td>
</tr>
<tr>
<td>theft, robbery, or fraud (national survey)</td>
<td>a4 b3 c6</td>
</tr>
</tbody>
</table>

Facets and elements

<table>
<thead>
<tr>
<th>Facets</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: injury</td>
<td>B: environment</td>
</tr>
<tr>
<td>a1: material</td>
<td>b1: primary</td>
</tr>
<tr>
<td>a2: psychological</td>
<td>b2: secondary</td>
</tr>
<tr>
<td>a3: unspecified</td>
<td>b3: unspecified</td>
</tr>
<tr>
<td>a4: unspecified</td>
<td>c1: absence</td>
</tr>
<tr>
<td>a5: physical</td>
<td>c2: work</td>
</tr>
<tr>
<td>a6: psychological</td>
<td>c3: economy</td>
</tr>
<tr>
<td>a7: unspecified</td>
<td>c4: social</td>
</tr>
<tr>
<td>a8: unspecified</td>
<td>c5: residence</td>
</tr>
<tr>
<td>a9: unspecified</td>
<td>c6: unspecified</td>
</tr>
</tbody>
</table>
The matrices of item-intercorrelations from both studies were submitted to ordinal MDS. The results are displayed in figures 5 and 6, respectively. As can be seen, the three facets specified in our mapping sentence (figure 4) could be identified in the pilot study and in the national survey. Both, injury and life areas play polarizing roles. They are superimposed on one another, splitting the bi-dimensional space into wedge-like regions. Environment, on the other hand, pops up as a modulating facet, separating stressors to the primary environment (inner circle) from those to the secondary (in the periphery; see Bilsky, 1996, 1999, Bilsky & Wetzels, 1994, 1995, for more details). All in all, our findings correspond to the structure suggested by Levy and Guttman’s model of adjustive behaviour (see figure 3).

![Figure 5: Stressors of personal safety: 2-dimensional ordinal MDS (pilot study; Germany 1992; N = 213).](image)

![Figure 6: Stressors of personal safety: 2-dimensional ordinal MDS (national survey; Germany 1992; N = 11,116).](image)
Fear of Crime

What, then, are the advantages of this type of analysis? - From my point of view, at least the following aspects should be mentioned:

First, and most importantly, we were able to specify a common, theory-based frame of reference for comparing criminal and non-criminal stressors of well-being with regard to different standards (facets). Systematic comparisons based on such a common frame are necessary for judging the relative importance of crime in general, and of concrete criminal acts in particular. They do not only facilitate the identification of social problems, which have been underestimated in the past - whether with respect to the general public or to special groups, e.g. elderly people. They are also helpful in avoiding overreactions to problems that do exist, but are of minor importance.

For illustrative purposes, the amount of strain caused by different stressors, as assessed in our own research (Bilsky, 1996; Greve, 1998; Wetzels et al., 1995), is displayed in figure 7. Data were collected in 1992 from a representative German sample (N=11,116). Items are grouped, displaying stressors to the primary environment first and to the secondary next, with the two crime-related items at the end. Within groups, stressors are arranged according to life areas (cf. figure 6). Strain caused by the different stressors is reproduced as deviation from the overall mean. Of course, it may vary considerably, depending on the subjects under consideration; for our data, this could be shown both for the old and new Federal States (Bilsky, 1996), and for different age groups (Greve, 1998).

Figure 7: Strain caused by criminal and non-criminal stressors
(national survey; Germany 1992; N = 11,116).
One might, of course, ask whether it is not sufficient (and much easier), just to stay with single stressors as standards of comparison. This type of comparisons has been practiced repeatedly in past research. The answer is both a conceptual and a methodological one: Comparisons will remain theoretically vague as long as there is no clear rule of how to sample stressors to be compared. In fact, the FT-approach goes beyond concrete items and formally defines the respective population from which they are sampled. This aspect is of considerable importance when comparing findings from studies, which used different operationalizations. As long as there are no valid standards for judging their conceptual equivalence, a reasonable transfer of research findings is problematic. However, isolated findings, which do not contribute to a systematic accumulation of knowledge, are scientifically useless.

The necessity to pay close attention to the form of operationalization of a specific concept under study is also evident from our data. It is probably not by chance, that ‘crime’, as analysed in our pilot study, popped up in the secondary environment, while the more concrete criminal acts, presented in the national survey (assault; theft, robbery, and fraud), showed up in the primary environment (figures 5 and 6). This finding widely parallels the distinction between ‘concern about crime’ and ‘personal fear of crime’ (Bilsky, Pfeiffer & Wetzels, 1993; Skogan, 1993).

Concern does not show the defining characteristics of an attitude, i.e., relative stability over time and relevance with respect to the directionality of behaviour. Furthermore, it often refers to an abstract issue that is beyond the scope or personal experience and influence. Thus, it might best be understood as an opinion in terms of social psychology (Bergler & Six, 1972), which is based on second-hand information (from the media or other distant sources) and likely to result in stereotypical behaviour, if at all. Fear, on the other hand, is usually much more concrete as regards personal experience with and behaviour towards the respective stressors.

Levy and Guttman’s (1989) theory of adjustive behaviour is a general approach which has the advantage of integrating psychological concepts like well-being, strain and coping, that come into play when investigating fear of crime. Insofar it is suited to serve as a common frame within which to compare stressors of personal safety in general, including crime. It is true though that, in the present form, it is but a rudimentary system that has to be elaborated by taking more recent findings from stress and coping research into consideration. This applies to both the refinement of facets and their elements. However, this task is beyond the scope of the present paper.

FACETS OF FEAR OF CRIME

In order to study fear of crime (foc) in more detail, and to distinguish it from related concepts like concern, it is necessary to modify, extend, and complement the present approach as a second stage. More precisely, it is necessary to elaborate on
Fear of Crime

the notion of problematic conditions of life (figure 2). In addition, the modality of a person’s reaction whose personal safety is threatened by a crime-related stressor needs specification. I will outline the essential features of this task using facet design, as applied in our own research (Bilsky, 1993; Bilsky, Pfeiffer & Wetzels, 1993; Greve, 1998). The following facets resulted from theoretical considerations and from scrutinizing criminological studies. Of course, this is not to say that all of them have to be specified in every particular study. On other occasions, however, it may prove necessary to introduce additional or to modify existing facets, depending on the focus of the respective research. In any case, a common (meta-) theoretical frame to which to refer when planning research, and when analysing data is helpful in accumulating and integrating knowledge systematically, both within and beyond a concrete study. This applies even more when confronted with seemingly contradictory findings. The following comments briefly outline why the different facets are integrated in the overall mapping sentence on fear of crime, as displayed below.

With regard to the criminal act, distinguishing property from personal crime is supposed to be basic for elaborating on fear of crime. This distinction is well established in criminological literature (cf., Skogan, 1987); it points to different targets of a criminal act as well as to different costs incurred by the victim. Imagining a victim of assault (personal), robbery (personal and property), or theft (property crime) conveys an intuitive understanding of this distinction.

The examination of the costs incurred by a victim of crime suggests retaining injury, already known from our previous analysis, as another facet for specifying fear of crime (Fattah & Sacco, 1989). While theft, for example, is normally associated with the anticipation of material loss, robbery often results in both material loss and physical harm. In addition, psychological impairment of the victim may be observed in the latter case, too. Finally, assault or rape is likely to result in high physical as well as psychological costs. Other examples that are associated with these types of costs could be easily found.

Looking at a person’s reaction to criminal stressors opens a third perspective on fear of crime. From this perspective, differentiating between an affective, cognitive, and behavioural mode of behaviour is a conceptually useful and promising specification of fear reactions (Bilsky, 1993; Greve, 1998). As in other domains of research (e.g., in attitudinal research; McGuire, 1983), reactions belonging to different modalities may either co-vary or differ substantially, depending on the respective situation. Consequently, this conceptual distinction is of considerable importance with respect to the choice of adequate measurement procedures.

Concentrating on the victim and the criminal act without paying attention to the offender and the situational context would certainly neglect the dynamic character of crime. Although, according to lay concepts of delinquency, crime is likely to be specified as an illegal outdoor activity of a stranger, this is only part of the truth. As known from criminological research, and confirmed by our own findings
FEAR OF CRIME, PERSONAL SAFETY AND WELL-BEING: A COMMON FRAME OF REFERENCE

(Wetzels et al., 1995; Wetzels & Bilsky, 1997), there are many different forms of intra-family violence, for instance, that clearly conform to legal definitions of crime. Nevertheless, these acts are very often labelled differently, and relatives, close friends, and family members are rarely called criminals in everyday language.

Finally, most investigations in this domain of research relate to anticipated or fictitious victimization. However, there have been other studies that investigate the impact of previous victimization on fear of crime, too. In our own research, for instance, victims older than 60 years of age reported a drastic increase in fear as a consequence of a former victimization (Bilsky & Wetzel, 1997). This finding may best be interpreted in terms of a heightened salience of vulnerability and a lack of coping resources in the sub-population of elderly victims (Greve, Hosser & Wetzels, 1996). It seems reasonable, therefore, to pay close attention to the effect of personal experience when planning a comprehensive study on fear of crime.

These as well as related considerations are summarized in the following mapping sentence of fear of crime (figure 8). This sentence formally defines the population of fear of crime items as the Cartesian product of the elements of facets A (modality) to G (injury). As a result of this, the range of validity of findings based on this definition, is both outlined and limited in a transparent and controllable way. This definition served as the formal basis for constructing the foc-items of our survey instrument (Bilsky, Pfeiffer & Wetzels, 1993). It clearly deviates from the definition of concern items, presented in figure 9 for the purpose of comparison. This type of items has been frequently used in opinion surveys, focusing on whether crime is perceived as increasing or decreasing, or whether it is supposed to be a more serious problem in one region as compared to another (Skogan, 1993). While concern items have been applied in the KFN-Survey, too, they will not be discussed in the present context.

<table>
<thead>
<tr>
<th>A: modality</th>
<th>B: experience</th>
<th>C: criminal act</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a₁: cognitively)</td>
<td>(b₁: anticipated)</td>
<td>(c₁: personal)</td>
</tr>
<tr>
<td>(a₂: emotionally)</td>
<td>(b₂: experienced)</td>
<td>(c₂: personal and property)</td>
</tr>
<tr>
<td>(a₃: behaviourally)</td>
<td>(b₃: unspecified)</td>
<td>(c₃: property)</td>
</tr>
<tr>
<td>(a₄: unspecified)</td>
<td></td>
<td>(c₄: unspecified)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D: offender</th>
<th>E: place</th>
<th>F: time</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d₁: stranger/s)</td>
<td>(e₁: at home)</td>
<td>(f₁: during the day)</td>
</tr>
<tr>
<td>(d₂: acquaintances/s)</td>
<td>(e₂: at work)</td>
<td>(f₂: in the evening)</td>
</tr>
<tr>
<td>(d₃: relatives)</td>
<td>(e₃: outside)</td>
<td>(f₃: at night)</td>
</tr>
<tr>
<td>(d₄: unspecified)</td>
<td>(e₄: unspecified)</td>
<td>(f₄: unspecified)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G: injury</th>
<th>Rₜₗᵣ</th>
</tr>
</thead>
<tbody>
<tr>
<td>(g₁: material)</td>
<td>(high)</td>
</tr>
<tr>
<td>and likely to result in (g₂: physical)</td>
<td>injury to self</td>
</tr>
<tr>
<td>(g₃: psychological)</td>
<td></td>
</tr>
<tr>
<td>(g₄: unspecified)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8: Mapping sentence of fear of crime.
The systematic differentiation proposed in the above mapping sentence (figure 8) is certainly helpful when describing the research field under consideration in some detail. But specifying fear of crime in terms of facets clearly goes beyond a mere verbal definition. This can be demonstrated again by means of data from the KFN-Survey (Bilsky & Wetzels, 1997; Greve, 1998), as shown in the following two examples.

The first one is supposed to illustrate both the complexity of the phenomenon under study, and its systematic decomposition by means of the facet approach. Two sets of four foc-items each, differing with respect to modality (Facet A) and criminal act (Facet C), were submitted to correlational analysis, together with the so-called standard item. This item has frequently been used in criminological research, in spite of considerable critique in the past (Boers, 1991; Fattah, 1993; Skogan, 1993). Among other things, this critique aims at the doubtful validity of the standard item. All nine items refer to anticipated crime (Facet B).

With respect to the two item sets, the following additional specifications can be made: Different offenders, and different places where crime might occur, as specified by Facets D and E, were explicitly mentioned but not varied in these items. Time of crime (Facet F) and injury (Facet G) remained unconsidered in the present context. The following sample item illustrates the type of assessment applied:

*If you think about yourself, how often are you afraid [Facet A], that you could, for example, be burgled, by someone you know or a stranger [Facet D], whether at home or elsewhere [Facet E]?* ...
*I am afraid ... of being burgled [Facet D]
(very frequently, frequently, sometimes, seldom, never)*
The complete facetization of the items (i.e., the structuples) and the results of the correlational analysis, including the standard item, are presented in Table 2 in the form of a multi-trait-multi-method (MTMM) matrix.

Table 2: Intercorrelations of fear indicators: national German sample (old Federal States; N = 3,631).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Afraid (freq. of being...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>burglled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beaten and injured</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mugged and robbed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sexually abused</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of being...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>burglled</td>
<td>(a&gt;b&gt;c&lt;d&lt;e)</td>
<td>0.59</td>
<td>0.46</td>
<td>0.53</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beaten and injured</td>
<td>(a&gt;b&gt;c&lt;d&lt;e)</td>
<td>0.41</td>
<td>0.69</td>
<td>0.51</td>
<td>0.34</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mugged and robbed</td>
<td>(a&gt;b&gt;c&lt;d&lt;e)</td>
<td>0.49</td>
<td>0.51</td>
<td>0.80</td>
<td>0.35</td>
<td>0.66</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sexually abused</td>
<td>(a&gt;b&gt;c&lt;d&lt;e)</td>
<td>0.24</td>
<td>0.35</td>
<td>0.36</td>
<td>0.65</td>
<td>0.34</td>
<td>0.45</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>How safe do you feel if you were alone, outside, at night in this area? (standard item)</td>
<td></td>
<td>0.43</td>
<td>0.40</td>
<td>0.47</td>
<td>0.44</td>
<td>0.35</td>
<td>0.32</td>
<td>0.37</td>
<td>0.35</td>
</tr>
</tbody>
</table>

As can be seen, correlations vary considerably. They range from r=.24 to r=.75 between items that relate to the cognitive (risk assessment) and the affective component (frequency of being afraid) of fear of crime. Furthermore, correlations between these indicators and the standard item proved to be only moderate (r=.32 to r=.47). The variability found between the above correlations, thus, underscores the complexity of the phenomenon under study.

However, this complexity can be disentangled by referring to facets A and C which summarize the distinguishing features of the two item sets in Table 2. To prove this, an ordinal MDS was applied to the correlations of the facetted items. As suggested by their structuples, the eight items were split by the MDS according to both facets. The resulting two-dimensional solution (coefficient of alienation $K = .04$) took the form of a grid (duplex), which separates the affective and the cognitive elements of Facet A on one dimension. Facet C shows up on the other one, with personal, personal and property, and property crime, as ordered elements. Given this distinction, the question arises, of whether these differences prove relevant when analysing the relation between the above foc-indicators and other variables.

An answer to this question is given by the second example that draws on a study of Greve (1998). He recently dealt with a problem that has plagued criminologists for years - the fear-victimization paradox (Eve, 1985; Fattah, 1993; Sacco, 1990). According to this paradox, research has repeatedly shown that women and elderly people exhibit higher levels of fear in spite of a supposedly lower risk of victimization as compared with other subpopulations. Greve (1998) analysed this...
Fear of Crime

phenomenon in considerable detail, showing convincingly that this ‘paradox’ can be resolved by differentiating cognitive, affective, and behavioural aspects of fear. I will describe his findings to the extent that they relate to the above facetization of foc-indicators.

Starting from the question of whether older individuals exhibit more fear of crime than younger ones, Greve (1998), using the standard question above (table 2), checked as a first stage, whether he was able to replicate the often cited finding that fear rises with age. The positive answer found to this question justified his assumption of building on data similar to those cited in the literature in support of the fear-victimization paradox.

Next, he verified whether the same results were obtained when referring to the affective fear component for assessment. For this purpose, items measuring the frequency of the four criminal experiences outlined in table 2 were aggregated to form an overall indicator of this component. In terms of the above mapping, the distinctive features of this global indicator of affect are characterized by \( \{a_2, c_4\} \). According to Greve, this measure is closely linked to the intensity of fear as measured by the item-format ‘How great is your fear of being ...’ (\( r = .81; \) cf. Greve, 1998, p. 301). All in all, his analyses with this indicator do not confirm the covariation of foc and age supposed by the fear-victimization paradox - whether based on mean or individual scores.

As a third stage, a general indicator for the cognitive fear component, \( \{a_4, c_4\} \), was constructed in a similar way by aggregating the respective items on the subjective probability of becoming a victim (table 2, above). Again, no age-related increase on this dimension of fear was found. Instead, the subjective probability of victimization even seemed to drop somewhat in the age groups over sixty years.

Finally, Greve computed an indicator of the conative fear component, \( \{a_3, c_4\} \), by aggregating eight different precautionary behaviours. Two typical examples of these items are: ‘I only leave the house after nightfall when it is absolutely necessary’ and ‘I avoid certain streets, squares or parks’ (cf. Greve, 1998; Bilsky, Pfeiffer & Wetzels, 1992). Contrary to findings relating to the affective and the cognitive fear component, the age curve of the conative indicator showed an increase, indicating that older persons behave more defensively and carefully than younger ones. The author concludes this part of his analysis by stating: “This finding proves illuminating with regard to the discernible increase in fear of crime observed using the standard question. Apparently, this way of assessing fear of crime addresses a situation which is closely related to avoidance behaviours ..., and those aspects obviously gain in significance with age” (Greve, 1998, p.291). Greve’s results are revealing with respect to common shortcomings in criminological research on the fear-victimization paradox in general, and on age and fear of crime in particular. They confirm that the latter is not a monolithic but a multifaceted construct. This is of special importance, because its components exhibit different associations with age, thus underscoring the necessity of distinguishing sub-concepts of fear of crime, as suggested by the mapping sentence outlined before.
Both examples emphasize the usefulness of a common frame of reference for deriving operationalizations from theoretical considerations, for disentangling confusing research results, and for integrating findings systematically. Of course, this is not to say that there is (only) one suitable way of how to define fear of crime, or even only one acceptable (set of) definition(s). Quite to the contrary, definitions have to be tailored in accordance with the research questions to be answered. Otherwise methodological artefacts are likely to confound results and to mask effects as has been shown elsewhere (Bilsky & Wetzel, 1997). With this task, however, a common frame of reference like Facet Theory is extremely helpful.

7  EPISODE

Fear of crime is an explosive issue that has often been discussed controversially, both in the general public and in politics. Usually, the media and politicians are blamed for willingly accepting research results that match their political interests or fit into the prevailing political climate and ignoring others. However, the exploitation of research that is supposed to serve one’s own interests is but one part of the problem. In fact we have to admit that, very often, there is a gap between theoretical and methodological know-how in social sciences and its use in criminological research. As a consequence, the ignorance of this know-how has contributed to the repetition and reiteration of equivocal and misleading research results. Since political initiatives and activities are quite likely to affect the situation of victims of (fear of) crime, it is a matter of scientific responsibility and ethics whether or not scientists stick to well established but obviously poor research practices. The often-indiscriminate use of data and their inadequate analysis that accompanied the spread of computers and the accessibility of complex programs coined the derogatory abbreviation GIGO, garbage in garbage out, more than three decades ago. It is up to us to design better studies, and to brand those that are deficient. To my understanding, better studies are needed - and possible - in this politically delicate and sensitive domain.

REFERENCES


Fear of Crime


INTRODUCTION

In the past 30 years, the fear of crime has come to be understood as a problem that can be independent of crime itself. Researchers exploring crime prevention and community safety have suggested that the success of any programme rests not only on the actual reduction of crime but also on public perceptions of crime. For example, results from opinion surveys conducted in several countries indicate that most people believe crime rates are rising, regardless of actual trends, which suggest a general decline in police-reported violent and non-violent crime over the previous five years (see Roberts, 2001).

The fear of crime is now considered to be a concept that is complex and multidimensional in nature (see Hale, 1996; Rountree, 1998). At the risk of oversimplification, for the purposes of this paper, fear of crime will be defined as being comprised of two broad, interacting concepts, namely, individual factors and community-based factors (e.g., Hough, 1995; Taylor & Hale, 1986; Vitelli & Endler, 1993).

Traditionally, factors associated with the individual (e.g., age, gender, ethnic origin) have been the focus of researchers exploring the fear of crime. This conventional approach, adopted primarily by researchers conducting large-scale national crime surveys such as the British Crime Survey and the General Social Survey (GSS) by Statistics Canada, have lead to many well-known conclusions. For example, researchers across many countries have found that, regardless of the measure used, women report higher levels of concern about criminal victimisation than do men (e.g., Ferraro, 1996; Mesch, 2000; Roberts, 2001; Smith & Torstensson, 1997; Smith, Torstensson, & Johansson, 2001). The economically disadvantaged, previous victims of crime, and those with physical vulnerabilities have also been identified as those who are consistently most likely to express a
fear of criminal danger (e.g., Carcach, Frampton, Thomas, & Cranich, 1995; Killias & Clerici, 2000; Kury, Obergfell-Fuchs, & Ferdinand, 2001; Smith & Torstensson, 1997).

More recently, researchers have begun to explore contributions beyond individual characteristics in explaining crime-related fears. For example, the importance of the “lived environment” and satisfaction with one’s community (i.e., community-based factors), have been noted as at least as important as individual factors in understanding the fear of crime (e.g., Fisher & Nasher, 1995; O’Mahoney & Quinn, 1999). Recently, Tulloch (2000) conducted a detailed analysis of people’s understanding of the fear of crime. She suggested that “reducing the fear of crime must move beyond identifying which groups are most afraid of crime, to addressing the different contexts and cues that induce unease, anxiety, and even fear” (p. 466). In comparison to the research attention and subsequent understanding of the relationship between individual characteristics and the fear of crime, less effort has been devoted to identifying key community-based factors.

2 | THE CURRENT PAPER

The overarching purpose of conducting the studies presented here was to explore aspects related to the fear of crime, as experienced by two communities, one in Western Canada and one in Western Australia. Localized crime audits were initially carried out with the purpose of providing these two communities with context-specific recommendations on how to reduce crime and fear of crime. In retrospect, after conducting these two surveys, a number of parallel community-based factors related to the fear of crime were identified. The purpose of the current paper was to highlight and discuss these parallel community-related variables as the primary considerations in reducing the fear of crime that might be applicable to a variety of communities. Namely, these factors were: a) the effect of the media on perceptions of crime; b) the importance of police relations with community residents; c) community-based crime prevention programmes; and finally; d) the influence of environmental conditions on perceptions of crime. The factors identified from these two initial studies will first be considered in terms of previous research. Following this, these factors will be (re)considered in light of the two localized crime surveys. The goal is to establish an understanding of the community-based factors that appear to be the most prevalent in affecting the fear of crime.

3 | COMMUNITY-BASED FACTORS REVISITED

3.1 | THE MEDIA

The media has been assigned a central role in fostering perceptions of an increasingly dangerous world. In fact, it has been noted that media news coverage is
the public’s primary source of information about crime (Warr, 2001). Most analyses of media crime reporting have been concerned with the potentially distorted perception created by the high proportion of reports of violent crime (e.g. Heath & Gilbert, 1996; Rosenbaum & Heath, 1990). In particular, it has been suggested that presentations of crimes that are local, random, and sensational are most likely to arouse fear (e.g., Chiricos, Eschholz, & Gertz, 1997; Heath & Gilbert, 1996; Stapel & Velthuijsen, 1996). As Bernard (1992) illustrated, the more the mass media report on crime, the more the readers become afraid of crime. Interestingly, Williams and Dickerson (1993) demonstrated differences in the amount of space and prominence given to crime, particularly violent crime, among 10 British daily newspapers. The researchers concluded that there was a positive relationship between newspaper reporting of crime and fear of crime among readers that appeared to be independent of the demographics of the readers (Williams & Dickerson, 1993). In a Canadian study, availability of media supplied crime experiences was found to be a significant predictor of general fear of crime for both males and females (Vitelli & Endler, 1993).

The mass media are thus a powerful amplifying mechanism when it comes to crime. Given that technology is developing and expanding worldwide, the influence of media on the perceptions of crime appears to be a factor that is globally applicable. Interestingly, however, few researchers have devoted attention to the potential benefits of using the media as a source of positive information as a means of decreasing the fear of crime.

3.2 | POLICE RELATIONS

It has been suggested that fear of crime may be significantly affected by perceptions of and experiences with the police (e.g., Sacco, 1998; Sacco & Nakhaie, 2001). Primarily, the possibility of a connection between public confidence in the police and fear of crime is important in terms of fear prevention policy. Using Canadian GSS data, Sprott and Doob (1997) concluded that respondents who reported a high level of fear were significantly more likely to hold negative views of the police and the courts.

There has been, however, a lack of convincing evidence in the literature supporting the importance of the perception of police effectiveness, as it has been reported that confidence in or satisfaction with police does not necessarily reduce the fear of crime overall (e.g., Bennett, 1991; 1994). Others have reported, however, that policing initiatives aimed at improving police and community relations have increased respondents’ involvement with crime prevention programmes and their overall satisfaction with police (Bennett, 1991; Kennedy & Veitch, 1997; Kobayashi & Suzuki, 1994). Given the theoretical notion that more police presence and a better relationship between police and community should influence fear of crime, it appears necessary to consider this issue as an important community-based factor that could widely affect the fear of crime.
Fear of Crime

3.3 | COMMUNITY-BASED CRIME PREVENTION PROGRAMMES

It has been suggested that even serious violent crime can be reduced in neighbourhoods of concentrated disadvantage when residents band together for informal social control (e.g., Sampson, Raudenbush, & Earls, 1997). Whether organized by the police department or by local residents, most community-based programmes posit that residents and police must work together to combat crime. These programmes can serve a number of purposes, including increasing citizen participation on key issues affecting the quality of life in communities (DeKeseredy & Schwartz, 1996). Some of the more visible programmes include Neighbourhood Watch or Block Watch programmes. These programmes are organized by the police and are designed to stimulate neighbourhood co-operatives that will actively watch for and report suspicious behaviour. The success of generic and generalized crime prevention programmes, however, is mixed at best (e.g., Brantingham & Brantingham, 1990; Bursik & Grasmick, 1993; Darian-Smith, 1993).

More recently, researchers have reviewed the benefits of community-based crime prevention programmes that have been developed to meet specific needs of a particular community. Notably, Donnelly and Kimble (1997) reported significant improvements in residents’ perceptions of crime after a community-wide effort to combat crime was implemented. The researchers reported that concern about crime decreased, as did the perceptions of the level of neighbourhood crime (Donnelly & Kimble, 1997). The notion that similar programmes specific to individual community needs seem to have an impact upon fear of crime is intriguing and requires further attention (Brantingham & Brantingham, 1990).

3.4 | ENVIRONMENTAL FACTORS

There is a considerable body of research available on the importance of the physical environment in relation to crime. Many researchers have found a positive relationship between the level of neighbourhood physical and social “incivilities” and fear of crime (e.g., Ferraro, 1995; LaGrange, Ferraro, & Supanic, 1992). In the “Signs of Incivility” theory (Lewis & Salem, 1986), researchers suggested that passers-by interpret declining physical characteristics of a neighbourhood as a devaluation of moral standards, which subsequently results in feelings of fear. Further, it has been suggested that socially and/or physically deteriorated areas offered better opportunities for crime (Bursik & Grasmick, 1993). Similarly, Newman (1972) suggested the “defensible space” theory, in which the assumption was made that crime and the fear of crime were affected by physical characteristics of community buildings. Both of these theories, though differing slightly, still appear to be supported in more recent literature. For example, in a recent

[1] Physical incivilities refer to trash, litter, graffiti; social incivilities refer to inconsiderate neighbours and vagrants.
Australian study, Borooah and Carcach (1997) concluded that lack of neighbour-
hood cohesion, neighbourhood incivility, and perception of relatively high neigh-
bourhood crime levels contribute significantly to the probability of being afraid
of crime and the risk of victimisation. In a recent Canadian study, researchers
suggested that improving the physical appearance of a community (e.g., remo-
ving garbage and graffiti) might reduce fear as much as reducing crime (Alvi,
Schwartz, DeKeseredy, & Maume, 2001). Relatedly, others have suggested chan-
ges to communities based on the “opportunity model”. Specifically, changing
street patterns and limiting entrances and exits to certain areas may quickly chan-
ge criminals’ choices of crime sites but may also more slowly build social cohe-
sion among residents (Donnelly & Kimble, 1997).

Other specific suggestions have included changes to lighting in higher crime
areas. Prior to the early 1990s, researchers found that improving street lighting
was successful in dramatically reducing criminal victimisation (e.g., Painter,
and European research on fear of crime, and noted that the “environmental
trend” was prominent, in that it was discovered that changing the environment
resulted in changes in the perception of crime, especially by women (e.g., Van
der Wurff & Van Wegen, 1989). More recently, in a study of students in Glasgow,
researchers found that improved lighting considerably lessened fears of victimi-
sation (Nair, McNair, & Ditton, 1997). As indicated by these researchers, there
exists the possibility that environmental factors may also be a universal conside-
ration related to the fear of crime.

4 | METHOD

Two cities of similar size and demographics were surveyed, including one city in
Western Australia (population 15,000) and one city in Saskatchewan Canada
(population 34,000). The city in Western Australia is a suburb of a larger center
and is, thus, very close in proximity to a much larger population of people. The
city in Western Canada, conversely, is considered more of a rural community,
with the closest city center being a 45-minute drive away. Both cities have their
own police force.

As many have suggested, researchers examining the fear of crime have been guil-
ty of adhering to an overly restrictive methodological framework, incorporating
only rigid questionnaires (e.g., Ditton et al., 1999; Hale, 1996; Farrall, Bannister,
Ditton, & Gilchrist, 1997). Thus, to best address issues of crime and safety, a
three-tiered approach to examining participants’ experiences was employed.
First, police-reported crime statistics were gathered from both cities. Second,
community perceptions of crime levels and fear of crime were assessed via a
randomly administered survey, constructed in consultation with community
representatives, employing both quantitative and qualitative response options.
Third, it was deemed imperative to conduct a number of interviews with various
community groups and individuals in order to gain a full understanding of the issues related to crime and fear of crime.

The study in Western Australia was conducted in 1999 and the study in Western Canada was conducted in 2001. The methodology for the survey component of each project was similar; surveys were hand-delivered and gathered the following day by research assistants. Each survey included items related to general perceptions of safety, of crime, and perceptions of their respective local police services. The surveys differed slightly in the presentation of items, but generally included the exact, or very similar items. The interview component of both projects differed somewhat, as only three interviews were conducted in the Australian project, whereas a more comprehensive group of community stakeholders (11 interviews in total) were included in the Western Canadian project. The slight differences in the two studies reflect changes made based on knowledge gathered from the first study (in Australia). Overall, however, the similarities have allowed us to make initial comparisons, while the unique features of both studies have served to enhance the results.

5 | RESULTS AND DISCUSSION

The results and ensuing discussion are constructed in the following fashion. First, objective crime statistics for both communities are briefly discussed to provide the reader with a sense of police-reported crime in each city. Second, the individual characteristics of the respondents to the survey are discussed briefly, demonstrating the comparability of the responses with previous research. Finally, in terms of the main purpose of the paper, the parallels between specific community-related variables associated with the fear of crime in both cities are explored.

5.1 | OBJECTIVE CRIME STATISTICS

Objective crime statistics were analyzed for selected offenses in both cities. In the Australian sample, crime trends were reviewed over a seven-year period (1992-1998). Total crimes against the person remained relatively stable; however, crimes against property increased in 1997/98, after a decrease in 1996/97, to almost meet the seven-year low experienced in 1991/92. The only other crimes that showed a consistent increase over the past seven years in this city were drug possession and vandalism/graffiti. Crime rates were similar to other areas in Western Australian for all offenses, except for offenses against property, which were somewhat higher. In the Western Canadian sample, crime trends were reviewed over a ten-year period (1991-2000). Crime trends were stable or had decreased for theft from, and of, motor vehicles, all types of assaults, possession of illegal drugs, and shoplifting. An increasing trend was noted for commercial break and enter. Overall, both communities had experienced a relatively stable crime rate, with minor fluctuations for specific crimes.
5.2 INDIVIDUAL CHARACTERISTICS

Demographics. A total of 854 participants completed questionnaires in both of the communities (314 participants in the Australian sample including 193 females and 116 males, [4 missing], and 540 participants in the Canadian sample, including 294 females and 246 males). The average age of respondents in both samples was 40-50 years of age. The average length of residence for participants in the Australian sample was 14 years, compared to 24 years in the Canadian sample.

General Perceptions of Safety. Responses to the items in this section were based on a seven-point scale (1 = “very unsafe”, 7 = “very safe”). First, participants from both studies indicated that they felt that there was a “moderate” crime rate in their community (overall $M = 3.92$), although participants in the Canadian sample ($M = 3.63$) reported a significantly lower perception of crime than those in the Australian sample ($M = 4.21$) ($t (847) = 7.10, p < .01$). Interestingly, in both studies, this perception did not significantly differ between men and women. Further, the vast majority of participants in both communities felt that crime rates had “become more of a problem” (Australia, 45.6%; Canada, 54.1%), or had “stayed about the same” (Australia 43.7%, Canada, 43.1%) since they had lived in the city. Fewer than 7% of both samples felt that crime had “become less of a problem”. Based on police-reported crime statistics (noted above), crime rates appeared to have remained stable over recent years, with the exception of crimes against property in both communities. Thus, individual perceptions of the crime rate increasing (based on police-reported crime only) may not be entirely accurate.

Although some participants felt that crime was increasing, it was evident from the responses that the majority of participants from both studies felt at least moderately safe in their communities (overall $M = 4.94$); however, once again, participants in the Canadian sample did feel significantly “safer” ($t (847) = 11.42, p < .01$). Not surprisingly, participants reported feeling “safer” in their homes and walking alone during the day than during the night (see Table 1).

<table>
<thead>
<tr>
<th>Table 1. Perceptions of safety</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>How safe would you feel at home alone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>during the day</td>
<td>6.35</td>
<td>6.42</td>
<td>5.31</td>
<td>5.01</td>
</tr>
<tr>
<td></td>
<td>5.75</td>
<td>5.16**</td>
<td>4.43</td>
<td>3.83**</td>
</tr>
<tr>
<td>How safe would you feel walking alone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>during the day</td>
<td>6.37</td>
<td>6.19</td>
<td>5.34</td>
<td>4.86*</td>
</tr>
<tr>
<td></td>
<td>5.01</td>
<td>4.24**</td>
<td>3.45</td>
<td>2.34**</td>
</tr>
<tr>
<td>How safe would you feel walking alone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in your neighbourhood during the day</td>
<td>6.59</td>
<td>6.48</td>
<td>5.55</td>
<td>5.14*</td>
</tr>
<tr>
<td></td>
<td>5.82</td>
<td>5.07**</td>
<td>3.73</td>
<td>2.63**</td>
</tr>
</tbody>
</table>

Note. 1 = “very unsafe”, 7 = “very safe”, *p < .01, **p < .001
It is also worth noting that, in common with the larger fear of crime surveys, significant differences emerged between the male and female respondents (see Table 1), even though their perception of crime rates did not significantly differ. There are a plethora of theories proposed by a variety of researchers attempting to explain this “fear-gender paradox”, but these are beyond the scope of this paper (see Mesch, 2000; Smith & Torstensson, 1997; Smith et al., 2001).

**General Perceptions of Crime.** In order to assess fear of crime in more depth, items were included to assess the degree to which participants felt a number of specific crimes were a problem in their community (see Table 2). Participants in both communities viewed vandalism and house breaking to be the more serious crimes. Although they differed on what they viewed as the least serious crime, sexual assault was commonly viewed as one of the least serious problems; however, women reported viewing sexual assault as a more serious problem than did men.

In a separate item, participants were asked if “youth crime” (committed by persons under the age of 18) was a significant problem, and participants from both groups strongly believed that it was (Australia = 81.7%, Canada = 89.4%). Further, participants were asked to identify the crimes that they most feared would personally affect them. Again, vandalism, followed by house breaking was noted by over 50% of participants in both samples to be the crimes that they feared the most. Thus, the crimes viewed as the most problematic were also those most feared.

<table>
<thead>
<tr>
<th>Canada</th>
<th></th>
<th>Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crimes perceived as <strong>most serious problems:</strong></td>
<td>Mean</td>
<td>Crimes perceived as <strong>most serious problems:</strong></td>
</tr>
<tr>
<td>Vandalism/Graffiti</td>
<td>5.45</td>
<td>House Breaking</td>
</tr>
<tr>
<td>Dangerous Driving</td>
<td>4.82</td>
<td>Vandalism/Graffiti</td>
</tr>
<tr>
<td>House Breaking</td>
<td>4.81</td>
<td>Robbery</td>
</tr>
<tr>
<td>Crimes perceived as <strong>least serious problems:</strong></td>
<td></td>
<td>Crimes perceived as <strong>least serious problems:</strong></td>
</tr>
<tr>
<td>Armed Robbery</td>
<td>2.95</td>
<td>Sexual Assault</td>
</tr>
<tr>
<td>Personal Fraud</td>
<td>3.07</td>
<td>Domestic Abuse</td>
</tr>
<tr>
<td>Sexual Assault</td>
<td>3.30</td>
<td>Drunken/Disorderly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Behaviour</td>
</tr>
</tbody>
</table>

*Note. Responses based on a 7-point scale, 1 = "not a problem", 7 = "serious problem".*

### 5.3 COMMUNITY-BASED FACTORS

The data thus far have been related to the participants’ individual characteristics and their perceptions of crime. We now turn to the four community-related factors that emerged as parallel in the two studies, including: a) the effect of the **media** on perceptions of crime; b) the importance of **police relations** with the community; c) the desire for **community-based crime prevention programmes**; and d) the influence of **environmental conditions** on perceptions of crime.
Participants in both groups were asked to provide information on the major source(s) of information used to form beliefs about the overall crime rate in their community (see Table 3). The purpose behind this question was to gain an understanding of the medium that participants believed was most influential in affecting their perceptions of crime. The media, primarily the newspaper, was noted by the vast majority of participants in both groups, to be the major source of information on which both their positive and negative beliefs about crime rates were based. Interestingly, participants noted that the lack of negative crime reporting that often occurred in their local paper served to increase their feelings of safety. To a lesser extent, rumours and gossip seemed to be another common source of information, as well as personal observation (i.e., witnessing crime). As has been previously noted (e.g., Heath & Gilbert, 1996; Warr, 2001), the data provided by these two samples reconfirms the importance of the media in influencing peoples’ perceptions about crime.

Table 3. Sources of information used to base belief about crime rates

<table>
<thead>
<tr>
<th></th>
<th>Canada</th>
<th>Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media</td>
<td>52%</td>
<td>34%</td>
</tr>
<tr>
<td>Talking</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>Observation</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>Know a crime victim</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>Having been a crime victim</td>
<td>8.3%</td>
<td>12%</td>
</tr>
</tbody>
</table>

In the qualitative component of the studies, participants were asked, “What do you think can be done to make people feel safer in your community”? First, educating and informing the public via the media was a common suggestion. Public education in the form of TV and radio was identified as potentially the most effective, as it was suggested that the majority of seniors and youth do not read the newspaper or brochures. In the Canadian sample in particular, it was suggested that a weekly report of crime statistics from the police should continue to be published in a local newsletter, as it was a very effective means of keeping the public up to date. This finding is congruent with researchers who have suggested that media source credibility is important (Koomen et al., 2000), and having the police provide updates of crime statistics appeared to be a trustworthy source of information to the people of the community.

Further, it was suggested that the newspaper (in particular) be used to promote the positive things that individuals (notably, youth) in the community are doing. Participants suggested that they often read about the negative things occurring in their community, primarily crimes committed by young offenders. Announcing and promoting the many positive things that young people do in their community, participants felt, may help to change the perceptions that youth are always the instigators of criminal activity and that young people are to be feared. Further, it was suggested that the media could be used more often as an outlet to educate people about how to protect themselves against crime, as well as how to get help if necessary.
It appears, based on previous research and on the data presented here, as though the media has a global influence on the perceptions of crime. Thus, it is perhaps one of the first avenues to be explored when attempting to reduce the fear of crime from a community perspective. It is understood that it is difficult, if not impossible to effect change in the global media coverage. Participants in both studies suggested, however, that they primarily based their beliefs about crime on their local newspaper, giving it much more credence than the more national or international news sources. Thus, working in co-operation with the media at a community level to report on positive aspects about the community, and relay information about crime prevention and safety, may be vital in reducing the fear of crime.

**Police Relations.** The second factor to be considered is the relationship between the community and the police. Participants in both samples were asked a number of questions related to their perceptions of the effectiveness of the police forces in their respective communities. First, participants were asked about their satisfaction with their respective police forces; both groups reported a moderate level of satisfaction (See Table 4).

Bivariate analyses were conducted to examine public confidence in the police and fear. Public confidence in the police was measured via responses to a question concerning whether the participants felt satisfied with the role of the police service in their communities. In both samples “satisfaction with the police service” was significantly correlated with “perception of safety” (Western Australia: $r = .18, p < .001$; Western Canada: $r = .16, p < .001$).

It is possible to strengthen the analysis by improving the reliability of the variables. This was completed by constructing “scales” made up from a number of related items with the reliability analysis statistic (Bennett, 1994). Specifically, the “Perceptions of Safety” scale was comprised of a combination of six items related to perceptions of safety when out during the day or night and when at home during the day or night. The scale “Satisfaction with Police” included three items related to satisfaction with police presence and police services. Once again, there is a significant correlation between satisfaction with the police and perceptions of safety in each community (see Table 5). These findings provide an initial indication (similar to Bennett, 1994) that there is a relationship between feelings of safety in one’s community and satisfaction with police.

<table>
<thead>
<tr>
<th></th>
<th>Satisfaction with Police in Dealing with Public Disorder</th>
<th>Overall Satisfaction with Police</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>2.88</td>
<td>3.71</td>
</tr>
<tr>
<td>Canada</td>
<td>4.21</td>
<td>4.38</td>
</tr>
</tbody>
</table>

*Note: 1 = “Very Unsatisfied”, 7 = “Very Satisfied”*
Based on interview responses, increased police presence was one of the most common suggestions put forth by participants. The vast majority of participants from both samples indicated that they would like to see more officers in the community day and night, patrolling in their cars, but also on foot and bike patrols. Residents desired more police-community interaction; they would like to see police officers provide more community talks and be present in school and community groups so that they are seen as “people” and not strictly enforcers. Participants suggested that hiring more police officers would allow the Service to play a larger “community policing” role.

It is interesting to note that although participants from both samples indicated at least a moderate satisfaction with their respective police forces, overwhelmingly, respondents indicated a strong desire for increased police presence. It appears as though this second factor related to community and police relationships should be one of the initial considerations when attempting to decrease the fear of crime. Although it was not possible with the current data to examine whether an increase in police presence would affect fear of crime, this will be an important avenue to explore in future research. Further studies from other demographic areas will also be warranted in order to gain a better understanding of the contribution of this factor in the reduction of community fear of crime.

Community-Based Crime Prevention Programmes. The third factor that emerged as important to both groups was the need for context-specific crime prevention programmes. Participants were asked about specific Crime Prevention and Community Safety Programmes offered by their communities. In both samples, the majority of respondents indicated that they were most aware of Neighbourhood Watch (in Western Australia) and Crime Stoppers (in Western Canada), which are two similar programmes that encourage community residents to be more vigilant of criminal behaviour and to report crime that they witness. Both groups indicated, however, that they did not feel the programmes were being run at the optimal level and in fact noted that they were not active in these programmes. In addition, there was the perception that there were very few alternative community-based programs related to crime and safety.

When asked specifically about which programmes would be the most beneficial to implement, participants indicated a desire to see more community-specific programmes, instead of the more generic Watch programmes. In line with previous researchers’ findings, the more generic programmes often fail (e.g.,

Table 5. Alphas for and correlations between the individual scales

<table>
<thead>
<tr>
<th></th>
<th>Canada</th>
<th>Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction with Police</td>
<td>.7388</td>
<td>.8151</td>
</tr>
<tr>
<td>Perception of Safety</td>
<td>.8638</td>
<td>.8830</td>
</tr>
<tr>
<td>$r$</td>
<td>•13**</td>
<td>•17**</td>
</tr>
</tbody>
</table>

Note: $r =$ Pearson Correlation, $\alpha = $ alpha, **$p < .001$
Fear of Crime

Brantingham & Brantingham, 1990; Darian-Smith, 1993); however, greater success has been reported with community specific initiatives (e.g., Donnelly & Kimble, 1997). So, although there appears to be an overall desire for the implementation of community programmes within both samples, participants indicated the importance of attending to the specific needs of individual communities. This suggestion is supported by recent conclusions highlighting the importance of local crime surveys (O’Mahoney & Quinn, 1999), indicating that although communities may appear similar, there are often marked differences that are important to consider (Fisher & Nasher, 1995).

Some specific suggestions offered by participants included the need for funding for prevention and education programmes, particularly involving school groups so as to provide youth with information before they are faced with the opportunities to participate in criminal activity. A number of social skills and parenting programmes were also suggested as imperative, as participants indicated that taking a primary preventative approach by educating parents might serve to decrease later criminal activity by youths. Another notable programme initiative in the Canadian sample was the “Citizen’s Police Academy”, which involved specialized training about the role of the police service. Individuals who had participated in this programme felt that it provided them with greater insight into and appreciation for the work that police officers in their community do. A number of people felt that this programme was very effective and indicated that they would like to see it expanded to reach more people in the community. Overall, this third factor of attending to community-specific crime prevention programmes appears to be a widely applicable factor and one of primary consideration when addressing the fear of crime in communities.

Environmental Factors. The fourth and final community-based factor was related to environmental conditions in one’s community. A majority of participants in both studies noted that vandalism/graffiti was the crime they viewed as most problematic in their communities, and was reported as a crime most feared. Parallel with previous literature, especially relating to the “incivilities theory”, people feel less safe when living in communities that appear vandalised or unkempt (e.g., Borooah & Carcach, 1997; Ferraro, 1995).

In response to the open-ended and qualitative questions, the majority of participants in both studies (over 80%) indicated that increased lighting in areas that were viewed as unsafe (e.g., mall parking lots), or where “teenagers” typically hung out, would be very beneficial in reducing fears of walking in those areas at night. Again, congruent with previous researchers (e.g., Nair et al., 1997; Vrij & Winkel, 1991), attention to lighting also seems to be an important core element present in both communities. In previous research, it was identified that a neighbourhood’s permeability also affects crime (Taylor & Gottfredson, 1986), which may help to explain why those in Western Australia report feeling less safe than those in the more isolated community surveyed in Western Canada. Thus, it might be important to consider street closures that could reduce non-resident
FEAR OF CRIME: AN INTERNATIONAL PERSPECTIVE OF COMMUNITY-BASED FACTORS

traffic, possibly decreasing opportunities for potential offenders who are “passing through” (see Donnelly & Kimble, 1997).

Although the importance of keeping a community “clean” has long been accepted and supported by researchers as an important means of reducing the fear of crime, we reiterate the importance here. We also highlight the importance of this factor as one that may be a pertinent consideration in most communities to focus on when attempting to reduce fear of crime.

6 | SUMMARY AND CONCLUSIONS

In the current discussion, the fear of crime was defined as being comprised of two complex, interacting components that contribute to its existence: characteristics of the individual and community-based factors. In terms of research attention and theoretical understanding, individual characteristics (e.g., age, sex) have been much more widely researched and accepted in terms of their contribution to the fear of crime. The community-related factors that contribute to the fear of crime have received less research attention overall, and thus, researchers have yet to compile the most significant factors that may influence perceptions related to the fear of crime. In this paper, the authors have (re)visited four community-based factors that may, in a number of different settings, contribute to the reduction of the fear of crime. The purpose of highlighting these factors is to provide an initial indication of the most influential areas to address when attempting to reduce the fear of crime from a community perspective.

After conducting two separate, localized crime audits in communities in different countries, four parallel community-related factors that appear to influence perceptions of the fear of crime were identified. Those factors were a) the effect of the media on perceptions of crime; b) the importance of police relations with the community; c) the desire for community-based crime prevention programmes; and d) the influence of environmental conditions on perceptions of crime.

The first factor discussed related to the influence of the media on the perceptions of crime, highlighting the results of previous researchers who have noted the tremendous impact that the media has in increasing the fear of crime (e.g., Chiricos et al., 1997; Heath & Gilbert, 1996; Warr, 2001). Results from the two localized crime surveys presented in this paper reconfirmed that the media is the primary source of information used in forming perceptions of crime. Previously, most analyses of media crime reporting have focused on the distorted perception created by the high proportion of violent crime reports (e.g., Rosenbaum & Heath, 1990). Although this finding is applicable to this research as well, participants in these studies further suggested that the media be used as an outlet to promote the positive aspects of their community. It is appreciated that globally, the media has a strong influence over the perceptions of crime and it is likely difficult or impossible to control what is reported; however, focusing on local, community-
Based media in reducing the fear of crime might be a very beneficial factor on which to focus.

The second factor discussed in this paper related to the importance of police and community interactions. In the current study, as in previous research (Bennett, 1994), satisfaction with the police force was significantly correlated with perceptions of safety. The vast majority of participants in both studies indicated that increased police presence would be a primary contributor in decreasing their fear of crime. In line with community-based policing initiatives that have been the focus of many policing organisations, it appears from this research that positive police and community interactions may be a very important factor to consider when aiming reducing the fear of crime.

The third factor that emerged reflected participants’ desire for community-specific crime prevention programmes. In both samples, participants indicated an awareness of the community “Watch” programmes, but also indicated a low level of participation in these generic programmes. As others have suggested, it appears as though the programmes designed specifically for communities, developed by querying their specific needs, are by far the more successful (e.g., Donnelly & Kimble, 1997). Participants in both studies generated specific recommendations related to their community. It appears as though involving community members in crime prevention programmes may be beneficial in reducing the fear of crime; it may also serve to promote community solidarity.

The final factor discussed was related to the importance of redressing problematic community environmental conditions. There has been a long history in the literature regarding the relationship between the importance of the physical environment and crime (e.g., Alvi et al., 2001; Borooah & Carcach, 1997; Lewis & Salem, 1986; Newman, 1972). Primarily, neighbourhood incivility is thought to contribute significantly to the fear of crime and to the risk of victimisation. In the current study, participants from both countries reported that the crime viewed as most problematic, as well as the crime most feared, was that of vandalism and graffiti. Focusing on controlling such incidents and immediately repairing deteriorating areas may be a very important step in the reduction of the fear of crime. Given this long-standing relationship between the physical environment and the fear of crime, this factor is also of particular importance in decreasing the fear of crime at the community level.

In conclusion, the researchers of this study are not suggesting that these community-related factors are necessarily applicable to all cities in all countries. In fact, the importance of attending to individual community needs it vital in reducing the fear of crime in specific areas. It is suggested, however, that based on the results of these two initial studies and the brief literature review, that these (and potentially other) community-based factors might be considered as the primary initiatives to reduce the fear of crime. We present this initial data and ensuing discussion as a first step in identifying the potential core elements that affect the fear of crime in a diversity of populations.
REFERENCES


Fear of Crime


Assessing measurement invariance in ‘fear of crime’

STEFAAN PLEYSIER, GEERT VERVAEKE, JOHAN GOETHALS & JAAK BILLIET

I | INTRODUCTION

As a legacy of the era when ‘fear of crime’ research originated, quantitative surveys are to the very day the dominant methodology for investigating fear of crime, “to the exclusion of virtually all others” (Ditton & Farrall, 2000). Ever since, this conservative tenacity to quantitative methodology has been questioned by numerous authors, mainly concentrating their criticism on the ‘number one measure’ of fear of crime: ‘how safe do or would you feel being out alone in your neighbourhood at night?’ (‘very safe/reasonably safe/somewhat unsafe/very unsafe’). Due to the persistence of this critique (rightly in our opinion), a tendency towards using scaling techniques, as a far better way of measuring a complex and multidimensional concept as ‘fear of crime’, arose. Choosing more complex measures however, gives rise to a number of ‘new’ issues related to measurement error, which have so far not been discussed in the ‘fear of crime’ literature. In this paper we would like to mention and exemplify some measurement problems, which are in particular (although not exclusively) related to fear of crime research: equivalence and invariance of ‘fear of crime’ scale-indicators for gender, age, education and social class. Although the initial interest is not substantive contextual outcome, as our aim is to illustrate how to assess measurement invariance in ‘fear of crime’ scales, we opted for substantive (not experimental) survey data. Using structural equation modelling, we analysed the fear of crime scale in the Flemish part of the 1999 General Election Study Belgium (N=1951) (Meersseman, Billiet, Depickere & Swyngedouw, 2001).
2 | OVERVIEW

Since the empirical discovery of ‘fear of crime’, originally as a ‘widespread public anxiety about crime’, in the US victim surveys of the ‘Commission on Law Enforcement and Administration of Justice’ (established by President Johnson in July 1965), literature and research rose almost exponentially. Notwithstanding the ‘thickness’ of this body of research, Ditton and Farrall conclude their review of the literature rather pessimistic to say the least; after almost 40 years of research into this topic “surprisingly little can be said conclusively about the fear of crime” (2000: xxi).

An important factor underlying and explaining at least partially the above conclusion is the conservatism inherent to the era – ‘the heyday of survey research’ – when ‘fear of crime’ research originated. Large-scale crime and victim surveys were developed and administered, mainly by telephone surveys, in order to have a more accurate estimate of the victimization rate and the ‘widespread public anxiety about crime’ (Block, 1993). As these large-scale surveys have a longitudinal design and therefore are administered on a recurrent base, conservatism not only points to the dominant methodology (survey) but within that methodology to the item level as well. The concept ‘fear of crime’ in these surveys is predominantly measured with a single, ‘standard’ item: ‘How safe do or would you feel being out alone in your neighbourhood at night?’ with answering categories: ‘very safe’, ‘reasonably safe’, ‘somewhat unsafe’, or ‘very unsafe’ (Ditton & Farrall, 2000; Hale, 1996; Ferraro & LaGrange, 1987). Recycling this item is symptomatic for the (conservative) choice of comparability (for large, national surveys: with the previous wave(s) and other nations; for small-scale surveys: with the large-scale survey), over reliability and validity. This choice disregards by large some fundamental methodological issues. Firstly, measuring attitudes or theoretical concepts (as ‘fear of crime’) with one single item ignores the complexity and in some cases multidimensionality of the concept. For obvious reasons this is a fundamental threat to the validity. Secondly, using only one item heightens the sensitivity to measurement errors such as question wording effects and interviewer effects, and therefore endangers the reliability of the ‘instrument’. As this ‘inappropriate operationalization’, as well as the ‘conceptual cloudiness’ that emanates from it – but also produces it –, are already repeatedly discussed elsewhere1, we confine ourselves to the consequences of and solutions to this critique.

3 | USING SCALES

Due to the above-mentioned critique, research today applies to an increasing extent scaling techniques in order to measure the concept ‘fear of crime’. Notwithstanding the fact that in essence scales are undoubtedly superior to single-item measures in order to capture attitudes or complex theoretical concepts, they give rise to a number of ‘new’ issues related to measurement error. Central to scaling techniques and factor analysis is the assumption that a number of observed (‘manifest’) items or indicators represent an underlying (‘latent’) factor or theoretical concept. It is then assumed that this latent construct is responsible for the variance between the observed indicators (covariance), and that, after extracting this common variance there is no systematic variance between the indicators (error covariance) left (Depickere, 2002).

As in the case of a single item measure, the evaluation of the constructed scale should discuss validity as well as reliability. Validity deals with the question whether a specific set of items covers the meaning of the construct we wish to measure, and is therefore more a consideration based on theoretical ground than on an empirical basis. For the same reason we will not deal with this aspect as such. Reliability on the other hand has to do with how consistent a scale is if it would repeatedly be used to measure the same latent construct. As already mentioned, factor analysis is based on the item intercorrelations of the indicators that construct the scale; measurement error, whether it is at random or systematic, can disturb the true correlation of these underlying indicators. More specific –and this is our focal point-, as correlations do not take into account within-sample differences between groups, differences in scale means might be due to true differences between groups on the underlying construct, or due to systematic biases in the way different groups respond to certain items (Steenkamp & Baumgartner, 1998; Vandenbarg & Lance, 2000; Welkenhuysen-Gybel & Billiet, 2002; Waeg, Billiet & Pleysier, 2000). This is in essence what ‘measurement invariance’ is all about.

“The general question of invariance of measurement is one of whether or not, under different conditions of observing and studying phenomena, measurements yield measures of the same attributes. If there is no evidence indicating presence or absence of measurement invariance – the usual case – or there is evidence that such invariance does not obtain, then the basis for drawing scientific inference is severely lacking: findings of differences between individuals and groups cannot be unambiguously interpreted” (Horn & McArdle, 1992:117; Vandenberg & Lance, 2000). As Horn and McArdle are very clear about the consequences of violating the assumptions of measurement invariance, these

[2] ‘New’ to criminology and more specific to the ‘fear of crime’ research; new techniques in assessing and controlling for measurement error in substantive research, originating from psychologists, psychometricians, sociologists, are introduced in criminological research only little by little.
assumptions are rarely tested. Neither criminological research in general, nor ‘fear of crime’ research in particular, is exception to this observation. This is disturbing for at least two reasons. First of all, a substantial part of the ‘fear of crime’ literature has empirically found and discussed different ‘fear of crime’ levels between different groups (gender, age, race, education, social class...). To our knowledge, measurement invariance was never tested, so differences in ‘fear of crime’ between groups are not necessarily a reflection of true differences, but possibly a reflection of different item interpretation between different groups in one or more of the ‘fear of crime’ indicators. Secondly, it is common methodological knowledge that lower educated or older respondents may have less clear views on certain types of items, especially in the case of more abstract items, balanced (positively and negatively worded) items within scales... (Billiet & McClendon, 2000). Ignoring this is a denial of the possible existence of measurement invariance.

Before discussing and illustrating how measurement invariance can be assessed in the context of a substantive survey, we briefly attend to the different forms of measurement invariance. In their attempt to assess measurement invariance in cross-national consumer research, Steenkamp and Baumgartner (1998) distinguished six different levels of measurement invariance. Configural invariance (‘weak factorial invariance’) is based on Thurstone’s simple structure principle and states that the indicators of a latent construct should have the same pattern of fixed and free factor loadings across different groups. Configural invariance does not indicate that different groups respond to the items in the same way, in the sense that obtained ratings can be meaningfully compared across groups (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). Metric invariance involves invariance of factor loadings for like items across groups and therefore is a stronger test of factorial invariance. Scalar invariance tests the assumption that cross-group differences in the means of the observed items are due to differences in the means of the underlying construct. Even if an item measures the latent variable with equivalent metrics across groups (metric invariance), scores on that item can still be systematically upward or downward biased. Comparisons of group means based on such additively biased items are meaningless unless this bias is removed from the data (Steenkamp & Baumgartner, 1998:80). Scalar invariance is in the case of ‘fear of crime’ scales a necessary but also sufficient guarantee for measurement invariance. Further tests of the other aspects of measurement invariance are therefore not necessary. However, Steenkamp and Baumgartner (1998) also mention factor covariance invariance (invariance imposed on the factor covariances), factor variance invariance (invariance imposed on the factor variances) and error variance invariance (invariance imposed on the amount of measurement error).

[3] Of specific relevance is the measurement of ‘fear of crime’ in large, longitudinal national and international crime and victim surveys. It is clear that cross-cultural and ‘cross-temporal’ measurement invariance are important issues to attend, as what is understood under ‘fear of crime’ can vary across nations due to cultural (linguistic) differences, but also across time within cultures.
4 | ASSESSING MEASURING INVARIANCE

4.1 | METHOD AND DATA

The data used for the illustration of the assessment of measurement invariance consists of the written part (drop-off questionnaire) of the 1999 General Election Study for Flanders (N=1951), collected after the Belgian General Election held on June 13th, 1999 (Meersseman, Billiet, Depickere & Swyngedouw, 2001). The ‘fear of crime’ (FOC) scale in this study is a set of 4 ordinal items with a 7-point response scale going from 1 (‘never’) to 7 (‘always’). The items of this scale are presented in table 1. ‘Gender’ is the original dichotomous variable, and ‘age’ (AGE) is the original metric (in years) variable. ‘Education’ (EDUC5) is a preconstructed ordinal variable with 5 categories (‘lower’, ‘lower secondary’, ‘higher secondary’, ‘higher’ and ‘university’). ‘Social class’ is a latent construct with 3 indicators; again ‘education’ (EDUC5), ‘income’ (INC3) as a preconstructed ordinal variable with 3 categories (‘-60,000 BF’, ‘60,000-100,000 BF’ and ‘+100,000 BF’), and ‘subjective social class’ (‘subjclass’), as the respondent’s perception of the social class he/she belongs to (4 categories: ‘working class’, ‘lower middle class’, ‘upper middle class’ and ‘upper class').

<table>
<thead>
<tr>
<th>Item name</th>
<th>FOC item</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘outside’</td>
<td>I feel unsafe in the neighbourhood</td>
</tr>
<tr>
<td>‘limited’</td>
<td>The idea that something could happen to me makes me feel limited in my freedom of movement</td>
</tr>
<tr>
<td>‘robbed’</td>
<td>I’m afraid of being robbed in the streets</td>
</tr>
<tr>
<td>‘home’</td>
<td>At home I feel safe</td>
</tr>
</tbody>
</table>

The data/tabulations utilised in this publication were made available by the ISPO –Interuniversity Centre for Political Opinion Research, sponsored by the Federal Services for Technical, Cultural and Scientific Affairs. Jaak Billiet, Marc Swyngedouw, Astrid Depickere and Erik Meersseman originally collected the data. Neither the original collectors of the data nor the Centre bears any responsibility for the analysis or interpretations presented here.
In order to assess measurement invariance for ‘gender’, ‘age’, ‘education’ and ‘social class’, we use the technique of confirmatory factor analysis via structural equation modelling (LISREL 8.3). Contrary to factor analysis within classical statistical software, structural equation modelling allows us to take systematic measurement error into account. Depending on the measurement level of the ‘control variable’ we opt for the multiple group comparison option (in the case of ‘gender’) or a restricted factor analysis (in the case of ‘age’, ‘education’ and ‘social class’). Because we are dealing largely with ordinal variables, we used Weighted Least Squares (WLS) estimations, which are based on polychoric correlations and an estimate of its asymptotic covariance matrix.

In order to evaluate the fit of a model, the Root Mean Square Error of Approximation (RMSEA) is preferred over the Chi-square value as a measure of overall fit, because the Chi-square criterion will most likely reject models in the case of large samples. Therefore, in this contribution the Chi-square value will only be used as a relative measure to compare different models, not as an absolute value of the fit of a model. In the case of relative large samples –as in our case– it is common to use a measure of fit which takes into account the error of approximation in the population as well as the precision of the measure itself (Jöreskog & Sörbom, 1996; Depickere, 2002). The RMSEA is such a measure of ‘close fit’ of a model. Models with a RMSEA < .05 are considered acceptable (Billiet & McClendon, 2000).

4.2 | RESULTS

The general principle of assessing measurement invariance, whether a multiple group comparison or a restricted factor analysis is used, is rather straightforward. As a point of reference, the basic factor model, i.e. the ‘fear of crime’ (FOC) model, is used (see fig. 1). To test whether the assumption of measurement invariance in (one of) the indicators for FOC holds, a potential ‘violator’ is introduced (Oort, 1992; Billiet & Welkenhuysen-Gybels, 2001). This is a variable (i.e. ‘gender’, ‘age’, ‘education’ and ‘social class’) with respect to which one or more indicators of FOC may be biased. Therefore, we test another factor model, equivalent to the basic model with the exception that the factor loading of one or more of the FOC indicators on the potential violator now are set free to be estimated. If such a model provides a significantly better fit than the basic model, that indicator (or

Although WLS estimations should be used if some or all of the observed variables are ordinal or discrete, and ML estimations (Maximum Likelihood) in the case of normally distributed variables, the choice between the two is not as straightforward as it seems. Jöreskog and Sörborn state that ‘if the sample size is not sufficiently large to produce an accurate estimate of the asymptotic covariance matrix, it is probably better to use ML or GLS” (1989:223). As a ‘sufficiently large sample’ is a relative notion, we opted to estimate all our models both with WLS and ML. As both procedures lead to equivalent conclusions, we selected the WLS estimations (standardized solutions) for presentation.
indicators) is biased with respect to the violator in question. This is also called ‘differential item functioning’ and is by definition a violation of the assumptions of measurement invariance.

Figure 1 and table 1 show that of the FOC scale the first three indicators are positively worded items, and the last item (‘home’) is negatively worded. Therefore, a negative factor loading on this indicator is expected and confirmed. However, the reversal of the direction in which this last item was formulated has probably confused the respondents, what resulted in a remarkable lower factor loading. Within the scope of measurement invariance, this is all the more interesting as we expect that this confusion is not evenly spread over all respondents. When discussing the potential different item functioning under the violators ‘age’, ‘education’ and ‘social class’, we will pay extra attention to this item.

4.2.1 ‘Gender’ as a potential violator

In order to assess the assumption of measurement invariance between male and female respondents, we analyse the basic model above with the multiple group analysis in LISREL. In a first step, a model without any constrains is tested; all

[6] A reliability analysis of the FOC scale with SPSS also indicated the problematic character of item 4. The value of Cronbach’s alpha (as a measure of internal consistency of a scale) for the full scale (4 items) is however sufficiently high (779), considered that this value depends on the number of items in a scale. Further, the RMSEA indicates that the model fits the data rather well, and lower factor loadings can also contain relevant information (Waage, 1997). However, omitting item 4 from the scale would increase Cronbach’s alpha to .861.
parameters in this model are separately estimated for men and women and there-fore allowed to differ from each other. This model is represented in figure 2.

Figure 2: FOC model for men (above) and women (below).

Chi - Square = 9.74, df = 4, P - value = 0.04500, RMSEA = 0.041

[7] To be clear, the parameters of men and women are estimated in one and the same proce-dure, resulting in only one model (although separate path diagrams); therefore the fit indi-ces (Chi-square, RMSEA...) are also identical for men and women.
As can be seen in figure 2, factor loadings do not differ all that much between men and women. This is however only a raw estimate of measurement invariance. In a next step we assess the measurement invariance by fixing the factor loadings between men and women. Under the assumption of measurement invariance it is expected that fixing factor loadings will result in a better model (not significantly worse compared to the model with free factor loadings (figure 2)). If this would not be the case, and factor loadings of the male respondents would be significantly different from those of the female respondents, this is clearly a violation of the measurement invariance assumption. In this case we could say that ‘gender’ is a violator resulting in different item functioning for the ‘fear of crime’ indicators.

Chi - Square = 13.38, df = 8, P - value = 0.09943, RMSEA = 0.028

Figure 3: FOC model with fixed factor loadings for men (above) and women (below).
Fear of Crime

Figure 3 indicates that the model with fixed factor loadings is a better model compared with the model with free factor loadings in figure 2. Therefore, we conclude that there is no reason to believe the FOC indicators are biased with respect to ‘gender’.

4.2.2. ‘Age’ as a potential violator

To detect different item functioning with ‘age’ as a possible violator, we use a second method, e.g. restricted factor analysis as an item bias detection method especially suited for continuous, interval or categorical data. Using this method has several advantages (Oort, 1992). Because it is not necessary to dichotomise or group item scores, this method makes maximal use of the available information. Also it does not need quite as large samples as in the case of multiple group comparison. Although the general principle of assessing measurement invariance, as mentioned above, holds for restricted factor analysis as well, strictly speaking it is not necessary to fit different models in order to detect different item functioning. If correctly specified, the output of an estimated restricted factor model in LISREL also indicates Modification Indices (MI). These MIs show how much the fit of the model will improve if the factor loading of a specific FOC indicator on the potential violator ‘age’ were set free to be estimated (Oort, 1992). Therefore, it suffices to inspect the MIs and the matching Expected Parameter Change (EPC), as an estimate of the expected change of the value of a fixed parameter when it is added to the original model as a free parameter. If the largest of all MIs is significant, the corresponding FOC indicator is prone to differential item functioning and should therefore be removed from the FOC scale. Oort adds however, “that it is recommended to remove an item with a significant MI only when it is associated with a substantively large EPC, be it positive or negative” (Oort, 1992:156). Figure 4 shows the FOC model with ‘age’ as a potential violator.

[8] Compared to the model in figure 2, the Chi-square in this model raises with 3.64 (df = 4). As this increase is not significant, the model with fixed factor loadings is not significantly worse compared to the model with free factor loadings. Doing not significantly worse with less information (fixing factor loadings), justifies our conclusion that we have obtained a better model. An evaluation of the RMSEA points in the same direction.

[9] If however a non-linear relation between the violator and the potential biased item is expected, restricted factor analysis is not the appropriate technique. As restricted factor analysis only applies to linear relations, it can only handle uniform bias (Oort, 1992).

[10] As a rule of thumb Oort suggest to calculate an Adjusted Critical value (AC) for the MI’s (Oort, 1992).

\[
AC = \sqrt{\frac{\chi^2}{(c + df - 1)}} \times C
\]

\(\chi^2\) refers to the overall \(\chi^2\) fit statistic of the model, \(c\) refers to the critical value for a chosen significance level in a \(\chi^2\)-distribution with 1 degree of freedom and \(df\) refers to the degrees of freedom of the model. Ac is then used as a decision rule, other than the statistical significance of the MI’s, to determine when a certain item is considered biased (if MI > Ac the item is considered biased). Also, it takes into account the sensitivity of \(\chi^2\) with respect to sample size as well as the fact that there is always some specification error in the model (Billiet & Welkenhuysen-Gybelis, 2001).
Before discussing possible violations of measurement invariance concerning age, we point to the fact that the above model is the first that fits insufficient according to the criterion we set out earlier (RMSEA > .05). Inspecting the LISREL output for the MI's and EPC's results in table 2.

**TABLE 2.** MI's and EPC's for the FOC indicators on 'age'

<table>
<thead>
<tr>
<th>FOC indicator</th>
<th>MI</th>
<th>EPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>outside</td>
<td>17.25</td>
<td>-0.07</td>
</tr>
<tr>
<td>limited</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>robbed</td>
<td>5.60</td>
<td>0.04</td>
</tr>
<tr>
<td>home</td>
<td>13.33</td>
<td>-0.08</td>
</tr>
</tbody>
</table>

The AC value for this table, with a 5% significance level, is 19.967. As none of the MI's exceeds this AC, the rule of thumb results in the conclusion that there is no different item functioning for age. If, for the sake of argument, we would ignore the AC criterion and follow the original LISREL output for the MI's and EPC's results in table 2.

Figure 4: FOC model with 'age' as a potential violator.

Chi - Square = 40.76, df = 5, P - value = 0.00000, RMSEA = 0.061
mean that older respondents tend to score lower on this item (‘feel safer’) than younger respondents with the same score on the underlying latent FOC construct. As for ‘home’, older respondents tend to score lower (‘feel unsafer’) than younger respondents with the same score on FOC. This would confirm our hypothesis that older respondents had more difficulties with the sudden reversal of this last item. The same could also hold for ‘outside’; as this is the first item with a 7-point response scale, following some more straightforward questions, older respondents could have more difficulties with this new ‘question style’. An alternative conclusion for the two items could be that older people have more difficulties interpreting vague or ‘formless fear of crime’ items; both ‘outside’ and ‘home’ are ‘formless fear of crime’ items and therefore possibly more disturbing for older respondents (Keane, 1992). Again, this is only done as an illustration, as strictly speaking, following Oort’s rule of thumb, there is no significant prove of different item functioning.

4.2.3. ‘Education’ as a potential violator

The same procedure is followed in order to detect different item functioning with ‘education’ as a potential violator. Figure 5 presents the fitted model and table 3 the MI’s and EPC’s

![Diagram]

Chi - Square = 24.47, df = 5, P - value = 0.00018, RMSEA = 0.045

Figure 5: FOC model with ‘education’ as a potential violator.
TABLE 3. MI’s and EPC’s for the FOC indicators on ‘education’

<table>
<thead>
<tr>
<th>FOC indicator</th>
<th>MI</th>
<th>EPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>outside</td>
<td>8.63</td>
<td>0.05</td>
</tr>
<tr>
<td>limited</td>
<td>0.34</td>
<td>0.01</td>
</tr>
<tr>
<td>robbed</td>
<td>7.10</td>
<td>-0.04</td>
</tr>
<tr>
<td>home</td>
<td>2.26</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The AC value for this table is 11.987; as the highest MI is lower than this AC value, ‘education’ is not a violator for the measurement invariance assumption.

4.2.4. ‘Social class’ as a potential violator

The model to test measurement invariance for ‘social class’ is slightly different; as already mentioned social class is not a manifest item but a latent construct with three indicators. The model in figure 6 fits sufficiently according to the rmsea criterion and the factor loadings for the indicators (‘subjective social class’ (subjclass), ‘income’ (inc3) and ‘education’ (educ5)) of the social class construct are also satisfactory.

Figure 6: FOC model with ‘social class’ as a potential violator.
TABLE 4. MI’s and EPC’s for the FOC indicators on ‘social class’

<table>
<thead>
<tr>
<th>FOC indicator</th>
<th>MI</th>
<th>EPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>outside</td>
<td>8.24</td>
<td>0.05</td>
</tr>
<tr>
<td>limited</td>
<td>0.48</td>
<td>0.01</td>
</tr>
<tr>
<td>robbed</td>
<td>7.30</td>
<td>-0.05</td>
</tr>
<tr>
<td>home</td>
<td>3.38</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 4 indicates that again there is no different item functioning involving social class as a potential violator. None of the MI’s in the table exceeds the AC value of 11.840.

5 | DISCUSSION

Although the substantive conclusion of this paper should be that there was no measurement invariance in the assessed ‘fear of crime’ measurement, this was not the purpose of our effort. Our aim was to indicate the importance of measurement invariance and to illustrate the assessment of measurement invariance of ‘fear of crime’ scales within substantive survey data. The approach to measurement invariance in this paper was almost purely a technical matter, concerning the reliability of measurement instruments. However, not checking measurement invariance or the reliability of survey instruments seriously undermines all substantive and contextual outcome and therefore threatens the validity of the research. Also, theoretical reflections and contextual knowledge is sometimes needed to interpret the results of what is in essence an issue of reliability. Therefore, reliability and validity are perhaps more inseparable than what is at first sight reflected in this contribution.

Assessing measurement invariance is one thing, answering the question what to do in the case of a violation of the measurement invariance assumption, is another thing. Multiple group comparison allows the detection of different factor loadings across different groups, but does not assist with how to deal with it. As for restricted factor analysis, Oort’s rule of thumb or recommendation to remove an item with a significant MI only when it is associated with a substantive large EPC, is not as helpful as one wishes. Leaving out a biased item in a scale with ten or more items needs less justification than in a four-item scale, as in our case. As there is no general or straightforward criterion, the researcher has to make an informed decision, based on the general factor model of the instrument, on the difference of the factor loadings between groups (multiple group comparison), on the significance of the MI’s and the value of the associated EPC’s...) and this, within a contextual and theoretical framework.

Notwithstanding our research interest in ‘fear of crime’, it should be obvious that the application of the techniques presented in this paper exceed the ‘fear of
crime' research. When using scales, measurement invariance is always an issue to be attended to. As there is a growing tendency within the field of ‘fear of crime’ or criminology in general towards using scaling techniques or structural equation modelling, we hope the issue and techniques presented in this contribution will gain attention and importance, as they already did within other fields of research.

REFERENCES


Fear of Crime


Fear of crime in transitional Poland

JANINA CZAPSKA & JOZEF WÓJCIKIEWICZ

I | INTRODUCTION

Systematic research of public opinion in Poland has shown that the growth in the crime rate following the political changes of 1989 brought with it an increase in people feeling threatened. Meanwhile, according to the Polish literature (Misiuk, 1998; Czapska, 1997), this feeling does not strictly correspond to the actual threat from crime, which is still lower than that in many of the countries of Western Europe, or other Central-Eastern European nations. The attitude commonly defined as “crime anxiety” is, in reality, a “conglomeration” of myriad fears and concerns, the causes of which lie not only in the escalation of crime, but in the overall social situation as well: the high unemployment rate, insecurity, worsening of the economic situation, and the corruption and inefficiency of State administration (Czapska & Kury, 2002; Kury, 2002).

Bearing in mind the complexity of the conditions and consequences that govern the phenomenon described as crime anxiety, it is possible to analyse the results from research into this phenomenon in Poland over the last 10 to 15 years. Studies systematically conducted by one of the largest public opinion research organisations – the Public Opinion Research Centre (CBOS) – indicate that, since the early 1990’s, a decided majority of those polled perceived Poland as a country that was dangerous to live in. The same studies also indicate, however, that safety in the actual place of residence received a decidedly more positive assessment. Close to two thirds of those polled (67%) described their place of residence as safe and peaceful. The reasons for such disproportion in the assessment of safety on a national level to that of safety in people’s own local vicinity can most often be discerned in the fact that opinions concerning safety on a macro (national) scale are formed not on a basis of personal experience, but are in large part dependant on information conveyed by the mass media, which present emphatic and explicit statements by politicians, and report on serious crimes committed with the use of violence (Widacki, 2002; Magiera, 1999). Despite
the fact that the most serious crimes are a statistically marginal phenomenon, they have a very strong impact on the imagination, leading to the formation of an image of Poland as a country that is unsafe. When assessing one’s own immediate vicinity, or area that’s known, familiar and directly experienced, the point of reference is not the perception of a general climate of danger, which one can develop under the influence of the media, but one’s own experiences, insights and observations, which encompass petty crime, disruption of peace and public order, or other incivilities.

One of the phenomena most symptomatic of the changes taking place in Poland, which also confirms a variety of sources for views concerning Poland and its immediate surroundings, is the level of change in these opinions following the transformation of 1989 (Table 1, 2).

**Table 1.** Assessment of Poland in terms of safety: “In your opinion, is Poland a country in which it is safe to live?” (Data by percentage)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>74</td>
<td>26</td>
<td>19</td>
<td>21</td>
<td>24</td>
<td>22</td>
<td>33</td>
<td>28</td>
<td>18</td>
<td>27</td>
</tr>
<tr>
<td>No</td>
<td>22</td>
<td>67</td>
<td>79</td>
<td>77</td>
<td>75</td>
<td>76</td>
<td>64</td>
<td>70</td>
<td>81</td>
<td>71</td>
</tr>
<tr>
<td>It’s hard to say</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

(Source: CBOS, 2002)

**Table 2.** Assessment of district, community, or village in terms of safety: “Would you call the place where you live safe and peaceful?” (Data by percentage)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>80</td>
<td>67</td>
<td>69</td>
<td>70</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>71</td>
<td>68</td>
<td>69</td>
</tr>
<tr>
<td>No</td>
<td>17</td>
<td>30</td>
<td>29</td>
<td>29</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>28</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>It’s hard to say</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

(Source: CBOS, 2002)

The proportions in the responses concerning Poland virtually underwent a reversal; where in 1987, 74% of Poles were of the opinion that they lived in a safe country, only 26% of them thought so in 1993. The proportions of individuals who espoused contrary views amounted to 22% and 67%, respectively. Opinions concerning safety in people’s place of residence worsened as well, although this change was considerably less substantial. According to a survey conducted in
October 2002 by the Centre of Public Opinion Research (OBOP, 2002), 56% of those surveyed feel safe at night in the vicinity of home, whereas 42% feel unsafe. Women perceive the level of safety on a national scale somewhat worse than men do. Residents of larger cities (over 100,000) are more critical than residents of rural areas and smaller towns (CBOS, 2002). The feeling that there is a threat in the immediate vicinity depends in large measure on the place of residence: it is felt by one in 6 people residing in rural areas or small towns, while in the largest cities one out of every two residents feels so. Since 1996, the CBOS has also been studying the subjectively perceived risk of victimisation (Table 3). The number of people claiming a strong fear of becoming the victim of a crime dropped over the past two years, which is particularly interesting in view of the fact that in the same time, the number of people who (according to their own words) fell victim to crime increased significantly. Similarly to the two previous instances, the risk of victimisation is felt primarily by inhabitants of large cities (100,000+ residents) – over 1/5 of those polled, whereas among residents of rural areas and smaller towns – less than one in ten.

It should be emphasised that the increase of the fear of crime noted in 2000 was of a particular nature, as it affected almost exclusively those who at the same time believed that Poland was an unsafe country (CBOS, 2000). Meanwhile, much of the heated political debate was directed against the courts for their alleged leniency towards criminals, and against the legislator. The support of the politic circles and media for the policy of law and order combined with the alarming news about further growth of the most dangerous crime had a stronger influence on the fear of crime than other factors, such as safety in the nearest environment.

### Table 3. Subjectively perceived risk of victimization: “Are you worried that you could become the victim of a crime?” (Data by percentage)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I’m not worried</td>
<td>30</td>
<td>37</td>
<td>35</td>
<td>40</td>
<td>31</td>
<td>34</td>
<td>37</td>
</tr>
<tr>
<td>I’m worried, but not that much</td>
<td>50</td>
<td>46</td>
<td>47</td>
<td>43</td>
<td>46</td>
<td>47</td>
<td>52</td>
</tr>
<tr>
<td>I’m very worried</td>
<td>17</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>21</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>It’s hard to say</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

(Source: CBOS, 2002)

The effect of fear of crime is, primarily, a decline in quality of life due to withdrawal from public life, avoidance of certain places, people, resignation from going out in the evening, voluntarily shutting oneself in one’s own house or dwelling – phenomena common to many societies. In recent years, video monitoring of public places such as streets, squares, public transit facilities, parks, etc. using cameras designed for that purpose has gained in popularity significantly.
Fear of Crime

Cameras are very often installed on the initiative of and with funding from regional government organs, and Poles in large part appraise these initiatives positively.

Another consequence of fear is the demand, voiced by both average citizens, as well as many politicians, for the strengthening of institutions of formal social supervision, primarily police forces, the justice administration, and remaining means at the disposal of criminal law. This demand most often takes the form of appeals for increasing the number of police officers and broadening their powers, e.g. in the area of firearms use (CBOS, 2002a), for designating an ever-growing number of behaviours perceived to be dangerous or undesirable as criminal, and for increasing the severity of administered punishments or making conditions for administering the penalty of incarceration more stringent (Krajewski, 2002). In a situation such as this, the problem of crime can easily become politicised, as citing the need to combat it is an easy and generally accessible means for politicians to gather public support. “Crime is becoming a very useful tool for political activity. Everyone knows that ‘something needs to be done’, and those to whom anti-crime messages are addressed will rarely complain of it.” (Czapska & Kury 2002, p. 10).

As emphasised earlier, the perception of a widespread threat may increase under the influence of various kinds of political moves, in which the fight against crime is utilised as an element in building a strong position for a given politician or political group. Hence, for example, the highest-on-record rise in Poles’ feeling threatened ever noted by the CBOS, in September 2001 – 81% of those polled then recognised Poland to be a dangerous country (Table 1) – was attributed to the initiation by the Justice Minister at the time, Lech Kaczyński, of a public campaign for the toughening of criminal law and criminal policy in conjunction with an enormous surge in crime (Widacki, 2002). Opinions concerning safety in Poland improved after the Minister, who had been referred to in the press as the “administrator of fear”, left office, despite the fact that the following surveys, in 2002, were conducted shortly after an incident in which a policeman died at the hands of gang members on the outskirts of Warsaw, something which in Poland remains (as yet) a rare occurrence, and could influence opinions on safety in the country.

Difficulties in fighting crime and strengthening the feeling of security through the means of community crime prevention are evident in Poland, although local government has become involved in crime prevention activities, particularly following the country’s administrative reforms of 1998. It seems, however, that a series of negative processes are a consequence of those inherited from the former, communist system: State control over institutions of social supervision, the destruction of social ties at various levels, and of a negative predisposition towards the Police (Czapska, 1999). It should be stressed that the attitude towards the Polish Police, despite fluctuations in the 1990’s, is currently positive. Despite very critical opinions concerning the effectiveness of the Police, its equipment, and...
FEAR OF CRIME IN TRANSITIONAL POLAND

competence of its personnel, it is serving society considerably better than the courts or the Parliament, in the opinion of the majority of citizens – in April, 2002, 55% of respondents thought so (in October, 2002 – even 68%), whereas 19% of Poles rated the lower chamber of the Parliament (Sejm) positively; and the courts – 22% (CBOS, 2002b; OBOP, 2002). The Police enjoys a considerable measure of trust as well – in March 2002, and in October 2002, 72% of Poles claimed to have trust in the Police, which was the best result since the force was formed, in 1990 (http://www.policja.pl/statys/ocena.htm; OBOP, 2002).

Among the many factors that can influence people’s feeling of security, the visible presence of police officers on the street and possession of firearms by citizens seem to play a particular role. In a document titled “Principal directions in the work of the Police in 2002”, the Chief of Police outlined two strategic directions of activity:

1. To bring about a tangible improvement in the public’s feeling of security in urban environments, with a specific emphasis on larger agglomerations, and maintain the high rating concerning the level of safety in rural areas and smaller towns,
2. Raising the level of safety in public places that are commonly considered dangerous due to committed offences and petty crime.

The first of these goals could be achieved by increasing both the number of police patrols on foot, and number of district constables, as well as instituting a substantive reassessment of the model and philosophy of their work, in accordance with the philosophy of community policing. Unfortunately, in the last few years, both of these spheres of police work have left much to be desired.

In 2000, police patrol officers constituted 31.5% of prevention forces; in other words, they numbered 18,329. In practice, due to various reasons, only about 7,500 served in daily patrols, the overwhelming majority of which were not foot patrols. Patrol officers constitute barely 19% of all police officers, which causes that the average number of residents for every one patrol officer is as much as 2,109, whereas in Saxon region of Germany, for example, it is only 450. Therefore, it is not surprising that 45% of the respondents (62% in the country; 23% in the cities with 500,000+ residents) do not perceive any foot patrols in the vicinity of home (OBOP, 2002). To make matters worse, the organisation of patrol services is inadequate: analyses are not conducted for the needs of the distribution of these services, the system of control and supervision over this service does not function sufficiently, and the patrol officers themselves, as well as their superiors, are unfamiliar with the basic rules of organisation and methods for performing patrol duties. Consequently, they lack basic tactical and technical knowledge as to the performance of this service (Paciorkowski, 2001). The patrol service cannot be guided by a “shepherd” philosophy according to which police officers, analogously to sheepdogs, guard the passive and defenceless citizenry against the wolves that are criminals. A police officer on patrol should fulfil at least three other significant functions:
Fear of Crime

1. Give attention to situations that are ostensibly minor, but have an adverse effect on inhabitants’ quality of life,
2. Render assistance to people in minor matters that arise in everyday life,
3. Increase the feeling of security in local communities (Czapska & Wójcikiewicz, 1999).

The results of surveys conducted between 2001 and 2002 among 4,556 respondents – inhabitants of a large voivodeship in western Poland – are unequivocal: as many as 41% of those surveyed feel threatened as a result of the inadequate number of police patrols, while only 9% feel so due to a tangible threat which they are either encountering or aware of! Up to half of the respondents, when asked about ways in which their feeling of security could be enhanced, postulated increasing the number of patrols (Kowalski, 2002).

Yet a service whose primary mission is the enhancement of the feeling of security is a service of district constables, police officers assigned to the same region for an extended period of time, realising (since 1998) the philosophy of community policing. Unfortunately, that service, despite a number of declarations, remains the Achilles’ heel of the Polish Police. It is regulated by directive no. 15 of the Chief of Police, from September 23rd 1999, concerning forms and methods of performance of duties by the district constable and district constabulary supervisor. According to § 5, a basic form of fulfilling the duties of a district constable is a tour that consists of making rounds within the boundaries of their beat and realising tasks consequent to their duties. The duration of these rounds should account for 60% of the district constable’s on-duty time, and his tasks, as specified in §§ 6-14, exhaust all conceivable obligations for a police officer – with one exception, though: the conducting of preparatory procedures (§ 12). The district constable in particular “initiates activities aimed towards resolving the problem of fear” (§ 11 pt 6). Practice, however, stands in sharp contrast to theory here.

One of the authors (J.W.) several years ago compared the district constable to the Yeti: everyone has heard about him, but very few have ever actually encountered him. The comparison made an unexpected career for itself, even beyond police journalist circles, and this succinct diagnosis unfortunately continues to hold true – 73% of the respondents do not know their district constable! (OBOP, 2002). One can become well acquainted with the real picture of this service thanks to several reports.

According to a survey conducted by the Police Central Command in 2000, around 50% (out of 10 thousand) of Polish district constables never went through any specialised training program. District constables did not conduct preparatory procedures in only 6 (out of 17) voivodeship Police headquarters (Podgórski, 2000). An audit done by the Chief Board of Supervision during the latter half of the year 2000 in 5 voivodeship headquarters yielded the following findings: in extreme cases, otherwise forbidden, the conducting of preparatory procedures
FEAR OF CRIME IN TRANSITIONAL POLAND

took as much as 70% of the district constables’ on-duty time, and their knowledge of the area in which they were serving was unsubstantial. District constables performed many tasks that did not fall within their range of responsibilities; most of their time was taken up by patrol duty, usually outside of their own beat!

From 40% to 65% of police officers surveyed replied affirmatively to the question: “Have you sought the position of district constable, and is it consistent with your predisposition and professional ambitions?” - but when the question was: “Would you change your professional position for another, with comparable wages?” as many as half of those surveyed gave an affirmative reply (Wieszczakowska, 2001).

Monitoring of 168 Police facilities (around 9%), which employed approximately 16 thousand police officers (16%), allowed for establishing that 83% of district constables were in fact working, with the remainder of them out on vacation, medical leave, etc. The working tenure as district constable in the case of 60% of functionaries did not exceed 5 years. There were very large fluctuations in terms of personnel. Half of the district constables (51%) worked in an area populated by 3 to 5 thousand inhabitants, one fourth had a region counting up to 3 thousand people, and the rest regions of 5 to 7 thousand (20%), 7 to 10 thousand (4%) and even over 10 thousand (1%). In 12 units (7%) there was one typewriter at the disposal of more than five district constables, and only in 23% of units did each one have their own typewriter. District constables had computers in only 18% of the units. They conducted preparatory procedures in only every fourth unit. In 17% of the units, district constables devoted more than half of their time on duty to administrative and office work. They complained about bureaucracy in every other unit (Cybulski, 2001).

The conclusions are obvious: district constables who are occupied with conducting preparatory procedures, buried under stacks of paperwork, and not present in the field can by no means be an influential factor in reducing fear of crime.

The situation in terms of the influence that citizens’ possession of firearms has on their sense of security is similar, albeit for other reasons. Polish law dating from 1919 advocates a restrictive policy in this area. The statute from May 21st 1999 (Journal of Laws No. 53, pos. 549) concerning weapons and ammunition, effective March 20th 2000, regulates the issue no differently. According to art. 10, the appropriate Police organ issues a permit for possession of a weapon if the circumstances cited by the applicant justify the issuing of such a permit. The criteria for such a decision are specified by jurisdiction of the Chief Administrative Court (NSA). Specifically, a permit for a gas weapon (which, incidentally, is now considered a firearm under the statute) can only be issued in the case where the danger to the life or health of the applicant is greater than that which the average citizen is exposed to (NSA verdict from March 9th 1994, III SA 702/93, OSP 1995, 7-8, pos. 157). In the opinion of the NSA, the general level of security, in the current legal conditions, does not justify widespread access to firearms ownership by citizens (NSA verdict from January 6th 1998, III SA 1365/96).
Fear of Crime

According to the Main Police Headquarters, by the end of 2001 as many as 749,606 permits for guns had been issued: among others 25,028 for handguns, 107,337 for hunting guns, 236,859 for gas weapons, 148,859 for airguns, and 748 for machineguns.

The stipulation to make firearms generally available is raised from time to time by politicians during election campaigns, as a supposed hallmark of a democratic society. However, it is characteristic of Polish society that it does not allow itself to be deluded by various types of “democrats”, and essentially does not desire the proliferation of firearms.

The results of two studies by the Public Opinion Research Centre from April 1996, and April 2001 (CBOS, 2001), are especially interesting (Table 4). Respondents were asked the same question: “In your opinion, in the current situation, should obtaining a permit to own a firearm be made easier for common citizens, or to the contrary – should laws concerning the issuing of firearms permits be made more stringent?” The answers were as follows:

Table 4. Firearms ownership: “In your opinion, in the current situation, should obtaining a permit to own a firearm be made easier for common citizens, or to the contrary – should laws concerning the issuing of firearms permits be made more stringent?”

<table>
<thead>
<tr>
<th></th>
<th>1996</th>
<th>2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firearms ownership</td>
<td></td>
<td></td>
</tr>
<tr>
<td>should be restricted</td>
<td>83%</td>
<td>68%</td>
</tr>
<tr>
<td>Firearms ownership</td>
<td></td>
<td></td>
</tr>
<tr>
<td>should be made easier</td>
<td>11%</td>
<td>20%</td>
</tr>
<tr>
<td>It’s hard to say</td>
<td>6%</td>
<td>12%</td>
</tr>
</tbody>
</table>

And thus, a decided majority of Poles continues to oppose the easing of access to firearms, while only one fifth of them advocate greater permissiveness in this area. Nevertheless, the latter have grown twofold in number over the last five years. “This is happening despite the fact that, in the same time, the number of those surveyed who have a strong feeling of being personally threatened has not changed, and the number of those indicating a threat from crime on a national scale and in the place of residence grew very insignificantly. This is due to the fact that attitudes regarding common citizens’ possessing firearms are neither linked to the perception of the threat level on a national scale, nor to that in the place of residence, and are tied only to a marginal extent with declarations of feeling personally threatened” (CBOS, 2001, p. 8).

Respondents were also asked about the possible consequences of firearms possession in the home:
TABLE 5. Consequences of firearms possession: “In your opinion, would a law enabling widespread possession of firearms in the home, without the possibility of carrying them in public, contribute to:

<table>
<thead>
<tr>
<th>Consequence</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limiting crime</td>
<td>36</td>
<td>56</td>
</tr>
<tr>
<td>Increasing the number of killings</td>
<td>66</td>
<td>27</td>
</tr>
<tr>
<td>Increasing the feeling of security</td>
<td>51</td>
<td>39</td>
</tr>
<tr>
<td>Increasing the number of firearms in the possession of criminals</td>
<td>78</td>
<td>16</td>
</tr>
<tr>
<td>Increasing the number of incidences of unjustified firearm use</td>
<td>83</td>
<td>9</td>
</tr>
</tbody>
</table>

Consequently, Polish citizens' opinions are generally consistent with the results of epidemiological studies on the consequences of firearms proliferation in other countries (e.g. Taubes, 1992; Mcdowall, Loftin & Wiersema, 1995).

In Poland, fear of crime is generally recognised as an independent, significant social problem, something that was expressed in the recognition of reducing fear as one of the primary goals of the national crime prevention program now being prepared, among other things. Without questioning the importance of this step (it is not yet known how the government intends to achieve this goal), it should be stressed that the politicising of this problem clearly influences Poles' views, adversely effecting opinions concerning security in the country and increasing punitiveness within society. The experience of the past 10 years in the area of action towards raising the feeling of security, despite numerous difficulties and failures, allows us to voice the conviction that this problem should be resolved primarily on a local level, with the rational and effective aid of the State.

The role of the Police in Poland in reducing the fear of crime is, unfortunately, negligible and will not change without increasing the number of district constables by 100%, and above all, introducing the philosophy of community policing into their daily practice. Considering that CP is more of a philosophy than a specific programme, it would be rather hard to assess its efficacy. However, we could point to several research projects on the application of CP, which corroborate the favourable influence of measures typical for CP, such as contact patrols, home visits, area-based foot patrols, promoting prevention in the form of multi-institutional partnerships, police, police stations accessible round the clock, emphasising the non-emergency function, etc. (Czapska & Wójcikiewicz, 1999).

Fortunately, the role of firearms in the reduction of fear is similarly negligible. Consequently, maintaining a restrictive policy in this area will not have an adverse effect on Poles' feeling of security.
REFERENCES


CBOS (2001), Poczucie bezpieczeñstwa obywateli i opinie o dostepie do broni palnej [Citizens’ feeling of safety and their opinions on access to firearms], Warszawa.

CBOS (2002), Poczucie zagrożenia przestępczoñci [Fear of crime], Warszawa.

CBOS (2002a), Opinie w sprawie zmiany przepisów o użyciu broni przez policjantów, [Opinions on the change of laws concerning the use of firearms by police officers], Warszawa.

CBOS (2002b), Oceny dziañania instytucji publicznych [Evaluations of the activity of public institutions], Warszawa.


CZAPSKA J., KURY H. (2002), WST_P [Preface], [in:] J. CZAPSKA, H. KURY (eds), Mit represyjnoñci albo o znaczeniu prewencji kryminalnej [The myth of repressiveness, or on the importance of crime prevention], Zakamycze, pp. 9 - 25.

CZAPSKA J., WÓJCIECHOWICZ J. (1999), Policja w spoñecznoñci obywatelskim [Police in civil society], Zakamycze.


KOWALSKI S. (2002), Poczucie bezpieczeñstwa w spoñecznoñci [The feeling of safety in the society], Gazeta Policjana, 28, 3.

KRAJEWSKI K. (2002), Punitiveness in spoilstwo polskiego [The punitiveness of Polish society], [in:] J. CZAPSKA, H. KURY, Mit represyjnoñci albo o znaczeniu pre-
wencji kryminalnej, [The myth of repressiveness, or on the importance of crime prevention], Zakamycze, pp. 173 - 197.

Kury H. (2002), Postawy punitywne i ich znaczenie [Punitive attitudes and their significance], [in:] J. Czapska, H. Kury (eds), Mit represyjno_ci albo o znaczeniu prewencji kryminalnej [The myth of repressiveness, or on the importance of crime prevention], Zakamycze, pp. 107 - 171.


Obop (2002), Poczucie bezpiecze_stwa Polaków i opinie na temat dzia_alno_ci Policji [Poles’ feeling of safety and their opinions on the activity of the Police], Warszawa.

Paciorkowski J. (2001), Znikaj_ce patrole [Disappearing patrols], Gazeta Policyjna, 22, 4.


Wieszczakowska H. (2001), Gorzej z praktyk_ [Things are worse in practice], Gazeta Policyjna 12, 1.


Widacki J. (2002), Polscy politycy o bezpiecze_stwie (mi_dzy demokracj_ a demagogi_) [Polish politicians on security (between democracy and demagogy)], [in:] J. Czapska, H. Kury (eds), Mit represyjno_ci albo o znaczeniu prewencji kryminalnej, [The myth of repressiveness or on the importance of crime prevention], Zakamycze, pp. 91 - 106.
PART 2
Witnesses
Laypeople’s criteria for the discrimination of reliable from non-reliable eyewitnesses

RAMÓN ARCE, FRANCISCA FARIÑA & DOLORES SEIJO

1 | INTRODUCTION

The models of information integration (e.g., Wigmore, 1935; Kaplan and Kemmerick, 1974; Kaplan, 1975; Kaplan, 1977; Kaplan, Steindorf, and Iervolino, 1978; Ostrom, Werner and Saks, 1978; Hastie, Penrod and Pennington, 1983) have identified two underlying dimensions of judgement-making i.e., reliability and validity. In the context of legal judgement-making, the former is generally associated to credibility of the testimony, and the latter to the weight of the evidence. Both dimensions have shown to be of great value and universal in the creation and prediction of narrative events in judgement-making by both experts and laypeople. In short, recent research (e.g., Arce, Fariña and Real, 2000) has revealed that reliability conferred to the different items of evidence mediates judgement-making in such a way that the evidence becomes polarised into two constellations i.e., one driven to guilt and the other to a not guilty verdict. This effect has been observed in jurors (Fariña, Fraga and Arce, 2000) and in judges (Arce et al, 2001). For example, in a case of rape the verdict frequently rests on the credibility assigned to the victim’s or aggressor’s testimony. Thus, if credibility is assigned to the victim, a guilty verdict will be reached or vice-versa. Moreover, gender has also found to mediate in the assignation of credibility in cases of rape (Arce, Fariña and Fraga, 2000). In other words, the assignation of credibility and validity of testimonies may bias judgement-making. Furthermore, research on metamemory and assignation of credibility underline that we overestimate our own and each other’s memory capabilities, and confer a considerable degree of reliability to eyewitness accounts, in particular those leading to a guilty verdict (McAllister and Bregman, 1986).
Witnesses

The two most prominent approaches to the evaluation of the credibility, reliability or perceptions of accuracy of judicial testimonies are based on legal and scientific criteria. According to legal doctrine, the reliability of a testimony is determined by the opportunity to observe (the ability to have observed the events, etc.), bias (control over possible interests), credit, temporal and inter-witnesses consistency, and plausibility (Wigmore, 1937; Schum, 1977; Hastie, Penrod and Pennington, 1983). From a scientific point of view, Wells and Lindsay (1983) confirmed by metanalysis that the inferences of accuracy included three types of information: variables to evaluate (phenomenology of the situation, encoding processes, and information retention), consistency (inter- and intra-testimonies), and biased responses (e.g., confidence). Likewise, two effective research tools for estimating the credibility of testimonies have been designed: one based on the verbal characteristics of the testimony (e.g., CBCA, and Reality Monitoring), the other relies on non-verbal behavioural analysis (see Vrij, 2000 for a review). In the assessment of credibility, however, people may resort to the use of subjective indicators, i.e., subjective models instead of the objective indicators i.e., objective models, reported in the literature, regardless of whether such behaviour is an accurate or reliable indicator of deception (Vrij, 2000). Furthermore, in judgement-making tasks, jurors have been found to be mediated in the assignment of credibility by variables such as attitudes towards the death penalty (Cowan, Thompson and Ellsworth, 1984), political attitudes, or gender (Fariña et al., 1999). Consequently, in order to estimate the consistency of a testimony three terms have been used in the literature: reliability, credibility and inferences of accuracy, according to whether the task refers to judgement making, evaluation of the credibility of a statement, and the perception of the accuracy of eyewitness identification. Nevertheless, these concepts share similar underlying mechanisms e.g., opportunity to observe, consistency in the statement).

Bearing in mind that the estimation of the credibility, reliability and the perception of accuracy of a testimony is a key point for judgement-making, this study aims to assess the criteria used by laypeople to assign reliability to the identification of two conflicting eyewitness testimonies on the bases of their statements. Moreover, we evaluate the scientific criteria for estimating credibility/reliability and the degree to which laypeople adequately employ the scientific criteria described in the literature, i.e., non-verbal behaviour and statement content analysis.

2 | METHOD

Subjects
A total of 200 subjects, mainly university students (70.4%) were selected, of which 34.7% were men and 65.3% women. The age ranged from 18 to 76 years with a mean age of 24.68 years and a standard deviation of 11.01.
LAYPEOPLE’S CRITERIA FOR THE DISCRIMINATION OF RELIABLE FROM NON-RELIABLE EYEWITNESSES

Material
The material used was a police video recording of two eyewitness accounts of a real-life case of robbery. The statements of the eyewitnesses were obtained using the cognitive interview procedure, and eyewitness identifications in the line-up were incompatible.

Procedure and design
The subjects were shown a police video of the conflicting statements of two real eyewitnesses, that is, they had identified two different suspects. Thereafter, subjects were asked to complete a questionnaire in which they had to:
- Decide which of the two testimonies was the most reliable, credible or accurate;
- Explain, in response to an open-ended question, why they believed one of the witness’s testimonies was reliable and the other was not;
- Independently evaluate the behaviour and the testimony of both witnesses using a questionnaire designed for this task (see “measurement variables”).

Measurement variables
Bearing in mind the objectives of this study, a battery of questions was designed to evaluate the following measurement variables:

1. First, using a one option only format, subjects had to select which of the two testimonies was reliable.

2. Thereafter, subjects underwent a “free narrative account interview” where subjects were required to motivate their reasons for considering one testimony as reliable and the disregarding the other. This interview was recorded on video.

3. The list of possible measurement variables of non-verbal behaviour indicative of deception is so extensive that only the measurement variables that have been reported in the literature as being objective and subjective indicators of lying were included in the study (Vrij, 2000). In order to ensure that these categories of analysis were present in the questionnaire, 10 experienced encoders classified the material into the 17 categories of the study using the procedures described by Thurstone. In this way, unproductive categories of analysis that were not detected in the material by at least one of the encoders were eliminated from the questionnaire. A total of 10 categories were found to be productive i.e., illustrator movements, smiling, gaze aversion, self-manipulations, high-pitched voice, speech hesitations, speech errors, speech rate, latency period, and hand and finger movements.

4. Content analysis criteria related to credibility i.e., C.B.C.A. were also evaluated (Steller and Köhnken, 1989). Similarly, 10 experienced encoders classified the 19 categories of the CBCA using the Thurstone procedure for classifying the material as productive or not. A total of 7 categories were found to be unproductive: accurately reported details misunderstood; pardoning the perpetrator; details
Witnesses

characteristic of the offence; related external associations; accounts of subjective mental state; raising doubts about one’s own testimony.

Analysis of protocols
Protocols related to motives of reliability or no reliability underwent content analysis in order to identify them. A reliable and valid mutually exclusive category system was designed, what Weick (1985) has labelled as a methodical category system. The categories were designed following the procedures outlined by Anguera (1990). The categories under analysis and their corresponding definitions were as follows:
• Gaze aversion: Looking at the face of the conversation partner;
• Length of the statement;
• Speech hesitations: Use of the words “ah”, “um” and so on;
• Lack of consistency with laws of nature;
• Structured account: Organised account of events;
• Clarity: Clarity, definiteness, specification, and vividness of the statement instead of dim and vague comments;
• Plausibility: Logical, coherent, and consistent account;
• Objective meta-statement: The witness describes what he saw and knows (facts), subordinating the interpretation to the facts;
• Authenticity: Degree to which the witness seems natural and relaxed;
• Confidence: The conviction in his/her testimony, being sure about what s/he saw, and gives a convincing testimony without showing signs of doubt;
• Conditional information: Exposure time, distance from the scene of the crime, was not distracted at that time;
• Contextual embedding: The witness’s ability to set an action in a specified time and space;
• Psychological processes: Witness’s ability to pay attention (focusing on the facts), perception (i.e., I saw, heard), and recall;
• Impression: Impression that the witness “seemed better”, “more sincere”, “more reasonable”;
• Statement focused on the facts: The testimony focused on the facts and events, without mentioning unrelated facts;
• Amount of details: The number of details provided by the witness’s testimony.

Reliability
Two encoders underwent exhaustive training prior to the encoding of the protocols, and were previously required to compare their encoding of protocols with other pattern codings that were not part of this study. The agreement index was taken as the contrastive procedure designed to correct bias in the pre-encoding stage. Moreover, one of the encoders had previous experience in other studies where the same procedure had been used (Jölluskin, 2000).

Thereafter, both encoders analysed the protocols (each encoding 50% of the sample responses). A week later each encoder encoded 10% of their own responses and 10% of the other encoder’s protocols.
In order to evaluate the within- and between-encoder consistency a concordance index was used with cut-off point of .80 (Tversky, 1977). The results obtained are shown in Table 1.

**Table 1. Within- and between-encoder consistency. Concordance index (CI)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>B-1/2</th>
<th>B-2/1</th>
<th>W-1</th>
<th>W-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychological processes</td>
<td>1.00</td>
<td>0.80</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Confidence</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Conditional information</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Amount of details</td>
<td>1.00</td>
<td>0.90</td>
<td>1.00</td>
<td>0.90</td>
</tr>
<tr>
<td>Speech hesitations</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Gaze aversion</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Clarity</td>
<td>0.80</td>
<td>0.90</td>
<td>1.00</td>
<td>0.90</td>
</tr>
<tr>
<td>Objective meta-statement</td>
<td>1.00</td>
<td>0.90</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Structured account</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Statement focused on the facts</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Contextual embedding</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Length of the statement</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Authenticity</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Lack of consistency with laws of nature</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Impression</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Plausibility</strong></td>
<td>1.00</td>
<td>1.00</td>
<td>0.90</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: CI = Agreements / (agreements + disagreements) X 100. B-1/2 = Concordance index between-encoders 1 and 2 from the material of encoder 1; B-2/1 = Concordance index between-encoders 2 and 1 from the material of encoder 2; W-1 = Concordance index within encoder 1; W-2 = Concordance index within encoder 2.

The results show that the data are within- and between-encoder consistent. Moreover, our encoders have also shown to be reliable in other studies (Jólluskin, 2000). Bearing in mind the within- and between-encoder consistency, the inter-studies consistency of the encoders, and the categorical system reliability, we can conclude that the input data is reliable (Wicker, 1975).
3 | RESULTS

3.1 | NON-VERBAL BEHAVIOUR (VRIJ, 2000)

The multivariate contrasts reveal that the reliability factor (“reliable witness vs. not reliable witness”) mediates significant differences in nonverbal behaviour, $F(10,261) = 7.350; p.001; T.E.=.220$. In other words, laypeople base reliability judgement-making on nonverbal behaviour.

Univariate analysis (see Table 2) show that lay assign credibility to those witnesses with fewer hand and finger movements, high speech rate, fewer smiles and speech hesitations and speech errors, with a low tone of voice, and a shorter latency period.

In short, laypeople link reliability, in line with the objective indicators, to a higher speech rate; fewer speech hesitations, fewer speech errors (repetitions of words and/or sentences, incompletions, slips of tongue); and fewer high-pitches of voice. Nevertheless, their interpretation may differ from the objective indicators (Vrij, 2000) by assigning value to reliability to control of non-functional hand and finger movements without moving arms, though the control of these movements is an objective indicator of unreliability. Likewise, smiling and laughing are not objectively evaluated by laypeople given that they are associated to little reliability, but objective models inform that there is no relation with the reliability of testimony. Moreover, illustrator movements, self-manipulations, gaze aversion, and latency periods are not considered by laypeople in assigning reliability; though this is not to say that these criteria are not taken into account in other contexts.

### Table 2. Univariate effects in non-verbal behaviour

<table>
<thead>
<tr>
<th>Variable</th>
<th>MS</th>
<th>$F$</th>
<th>$p$</th>
<th>$\text{Eta}^2$</th>
<th>$M_{\text{re}}$</th>
<th>$M_{\text{nre}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illustrators</td>
<td>.88</td>
<td>.22</td>
<td>.638</td>
<td>.001</td>
<td>4.64</td>
<td>4.53</td>
</tr>
<tr>
<td>Hand and finger</td>
<td>18.08</td>
<td>5.19</td>
<td>.023*</td>
<td>.019</td>
<td>3.1</td>
<td>3.62</td>
</tr>
<tr>
<td>Self-manipulations</td>
<td>.251</td>
<td>.08</td>
<td>.778</td>
<td>.000</td>
<td>4.01</td>
<td>4.08</td>
</tr>
<tr>
<td>Speech rate</td>
<td>124.77</td>
<td>52.73</td>
<td>.001***</td>
<td>.163</td>
<td>4.63</td>
<td>3.28</td>
</tr>
<tr>
<td>Smiling</td>
<td>19.46</td>
<td>7.96</td>
<td>.005**</td>
<td>.029</td>
<td>3.1</td>
<td>3.56</td>
</tr>
<tr>
<td>Speech hesitations</td>
<td>35.23</td>
<td>9.23</td>
<td>.003**</td>
<td>.033</td>
<td>3.63</td>
<td>4.35</td>
</tr>
<tr>
<td>Gaze aversion</td>
<td>6.252</td>
<td>1.88</td>
<td>.172</td>
<td>.007</td>
<td>3.22</td>
<td>3.53</td>
</tr>
<tr>
<td>Speech errors</td>
<td>36.34</td>
<td>13.01</td>
<td>.001***</td>
<td>.046</td>
<td>3.21</td>
<td>3.94</td>
</tr>
<tr>
<td>High-pitched voice</td>
<td>43.92</td>
<td>13.</td>
<td>.001***</td>
<td>.046</td>
<td>2.99</td>
<td>3.79</td>
</tr>
<tr>
<td>Latency period</td>
<td>4.47</td>
<td>2.2</td>
<td>.139</td>
<td>.008</td>
<td>2.35</td>
<td>2.61</td>
</tr>
</tbody>
</table>

**Note:** D.F.(1,271). ***$p .001$; **$p .01$; *$p .05$. $M_{\text{re}}$= Mean of reliable eyewitness; $M_{\text{nre}}$= Mean of not reliable eyewitness.
In short, nonverbal behaviour is a useful parameter for laypeople in estimating the reliability of testimonies. However, though it may be of value for assessing reliability, this does not imply that they are used correctly i.e., in line with objective models described in the scientific literature.

### 3.2 UNIVARIATE EFFECTS IN NON-VERBAL BEHAVIOUR

The existence of significant differences in the Content Analysis Based on Criteria (Steller and Köhnken, 1989) mediated by the reliability assignation factor (reliable witness vs. not reliable witness) was verified, $F(12, 331)= 7.465; p.001; T.E.=.213$.

The univariate effects (see Table 3) reveal that the testimony of reliable witnesses is perceived as being well set in time and space; makes more references to the perpetrator’s mental state; provides more detailed descriptions of people and places; describes in greater depth social interaction; and the accounts are logically structured. In contrast, the testimony of not reliable witnesses show more lacks of memory and spontaneous corrections.

### Table 3. Univariate effects in CBCA criteria

<table>
<thead>
<tr>
<th>Variable</th>
<th>MS</th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2$</th>
<th>$M_{re}$</th>
<th>$M_{nre}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical structure</td>
<td>100.67</td>
<td>37.85</td>
<td>.000***</td>
<td>.100</td>
<td>5.19</td>
<td>4.11</td>
</tr>
<tr>
<td>Amount of details</td>
<td>216.18</td>
<td>58.18</td>
<td>.000***</td>
<td>.145</td>
<td>4.89</td>
<td>3.3</td>
</tr>
<tr>
<td>Contextual embedding</td>
<td>104.11</td>
<td>38.32</td>
<td>.000***</td>
<td>.101</td>
<td>4.85</td>
<td>3.75</td>
</tr>
<tr>
<td>Descriptions of interactions</td>
<td>18.64</td>
<td>5.33</td>
<td>.022*</td>
<td>.015</td>
<td>3.47</td>
<td>3.01</td>
</tr>
<tr>
<td>Unexpected complications</td>
<td>10.43</td>
<td>2.79</td>
<td>.096</td>
<td>.008</td>
<td>3.68</td>
<td>3.33</td>
</tr>
<tr>
<td>Unusual details</td>
<td>.17</td>
<td>.05</td>
<td>.817</td>
<td>.000</td>
<td>3.3</td>
<td>3.25</td>
</tr>
<tr>
<td>Superfluous details</td>
<td>.01</td>
<td>.01</td>
<td>.981</td>
<td>.000</td>
<td>3.4</td>
<td>3.4</td>
</tr>
<tr>
<td>Perpetrator’s mental state</td>
<td>97.16</td>
<td>24.33</td>
<td>.000***</td>
<td>.066</td>
<td>3.13</td>
<td>4.19</td>
</tr>
<tr>
<td>Admitting lack of memory</td>
<td>.24</td>
<td>.08</td>
<td>.779</td>
<td>.000</td>
<td>2.52</td>
<td>2.47</td>
</tr>
</tbody>
</table>

Note: D.F.(1,343). ***$p .001$; **$p .01$; *$p .05$. $M_{re}$ =mean of reliable eyewitness; $M_{nre}$ =mean of not reliable eyewitness.

Briefly, Criteria Based Content Analysis is sensitive to the reliability assigned by laypeople, which coincides with empirical predictions. Nevertheless, the univariate effects indicate that not all of the variables were relevant and, contrary to the scientific predictions of the objective model, the admission of lack of memory
Witnesses

and spontaneous corrections are linked to an unreliable testimony. This lack of consistency with the empirical findings has also been reported in other studies (e.g., Biland, Py and Rimboud, 1999).

3.2 | CRITERIA BASED CONTENT ANALYSIS
(STELLER AND KÖHNKEN, 1989)

Table 4 shows the criteria mentioned by laypeople to an open-ended question concerning their reasons for assigning reliability or not. The data reveal that laypeople use multiple criteria including criteria pertaining to content analysis of the statement (e.g., amount of details, logical structure), non-verbal behaviour criteria (e.g. vocal characteristics and non-vocal), criteria (e.g., plausibility), scientific criteria (e.g., conditional information and psychological processes), meta-memory (confidence), testimony validity criteria (inconsistent with the laws of science).

Table 4. Lay criteria of reliability

<table>
<thead>
<tr>
<th>Variables</th>
<th>N (%)</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focused on the facts</td>
<td>3 (1.5%)</td>
<td>-2.27</td>
</tr>
<tr>
<td>Clarity</td>
<td>29 (14.5%)</td>
<td>6.17*</td>
</tr>
<tr>
<td>Amount of details</td>
<td>103 (51.5%)</td>
<td>30.19*</td>
</tr>
<tr>
<td>Structured account</td>
<td>2 (1%)</td>
<td>-2.6</td>
</tr>
<tr>
<td>Inconsistent with the laws of science</td>
<td>4 (2%)</td>
<td>-1.94</td>
</tr>
<tr>
<td>Length of the statement</td>
<td>1 (0.5%)</td>
<td>-2.92</td>
</tr>
<tr>
<td>Location in time and space</td>
<td>1 (0.5%)</td>
<td>-2.92</td>
</tr>
<tr>
<td>Speech hesitations</td>
<td>4 (2%)</td>
<td>-1.94</td>
</tr>
<tr>
<td>Gaze aversion</td>
<td>1 (0.5%)</td>
<td>-2.92</td>
</tr>
<tr>
<td>Impression</td>
<td>23 (11.5%)</td>
<td>4.22*</td>
</tr>
<tr>
<td>Objective meta-testimony</td>
<td>16 (8%)</td>
<td>1.94</td>
</tr>
<tr>
<td>Conditional information</td>
<td>44 (22%)</td>
<td>11.04*</td>
</tr>
<tr>
<td>Plausibility</td>
<td>23 (11.5%)</td>
<td>4.22*</td>
</tr>
<tr>
<td>Psychological processes</td>
<td>35 (17.5%)</td>
<td>8.12*</td>
</tr>
<tr>
<td>Confidence</td>
<td>60 (30%)</td>
<td>16.22*</td>
</tr>
<tr>
<td>Authenticity</td>
<td>45 (22.5%)</td>
<td>11.36*</td>
</tr>
</tbody>
</table>

Note: * the number of case observed is greater p<.001 than expected (.05).

Nevertheless, the fundamental criteria, that is, those who significantly exceed the critical value of 5% of the motives, reduce the number, of an initial total of 16, the next 8:
LAPEOPLE’S CRITERIA FOR THE DISCRIMINATION OF RELIABLE FROM NON-RELIABLE EYEWITNESSES

- Clarity: Clarity and vividness of the testimony (criteria of the content of the testimony);
- Confidence: the witness’s conviction during the testimony, that is, the witness is sure what s/he saw, and made a convincing testimony without doubting (scientific criteria of metamemory);
- Amount of details: quantity of details during the testimony (content criteria of the testimony);
- Impression: Impression that the eyewitness was “better”, “more sincere”, “more reasonable” (criteria of nonverbal behaviour referring to personality of the speaker);
- Conditional information: exposure time, distance from the crime scene, not distracted at the time of the crime (scientific criteria for estimating variables, legal criteria of opportunity);
- Plausibility: logical structure, coherence and internal consistency of the account (legal criteria of plausibility, content criteria of logical structure, scientific criteria of reconstructability of the story);
- Psychological processes: ability to pay attention (focused on the facts, distraction), perception (i.e., I saw, heard) and witness recall of events (scientific criteria for estimating variables);
- Authenticity: relaxed or tense witness (criteria of nonverbal behaviour centred on non-vocal characteristics).

It is worth noting that in free recall tasks, subjects remember the most prominent information and they do so in a scale of importance (Loftus et al., 1992), which mediates judgement-making. According to our findings laypeople in discrimination tasks of reliable or unreliable witnesses rest judgement-making on several criteria: legal, scientific, content analysis, and nonverbal behaviour. The objective value of the scientific, legal, content analysis and nonverbal behaviour criteria has been well documented in the literature. Moreover, most of the criteria are properly used i.e., in line with the predictions of objective models, notwithstanding some exceptions such as the witness’s confidence in his/her testimony that is interpreted as a sign of reliability. Yet, it is well known that confidence and accuracy are not synonymous. Similarly, the nonverbal behaviour criteria used by laypeople to discriminate are not considered by the literature to be objective indicators of reliability. In fact, the authenticity can lead to the “Othello Error” (Ekman and O’Sullivan, 1989), and the relationship between impressions of reliability and accuracy is imprecise (Riggio and Friedman, 1983). Furthermore, these strategies are closely associated to bias in judgement-making, which involves more than the implicit knowledge of laypeople about their own and others memory (Wells and Lindsay, 1983; Lindsay et al., 1986).

4 | DISCUSSION

It is worth pointing out that care should be taken concerning the generalisations inferred from the findings. First, this study performed on mock decision makers...
Witnesses who do not have to perform as real ones (Fariña et al., 1994). Second, the empirical models applied i.e., Non-Verbal Behaviour and CBCA, were specifically designed to measure lying and deception of a victim’s statement, and not eyewitness reliability in an identification task as was the purpose of this study. Third, the measures of the scientific criteria were drawn from a recognition task but this does not imply it is of value in the laypeople’s evaluation of eyewitness reliability. Fourth, given that laypeople and legal experts are said to perform the same task in the evaluation of the reliability of an eyewitnesses testimony, universality in the estimation of reliability can be assumed, which supposes that the data may be generalised to legal experts (Schooler, Gerhard and Loftus, 1986).

Bearing in mind these limitations, the results obtained enable us to draw the following conclusions:

• Laypeople are able to recognise with a degree of reliability the scientific criteria of nonverbal behaviour linked to the reliability of a testimony;
• Likewise, laypeople can recognise with a high degree of accuracy the objective interpretation of the content criteria outlined in the CBCA designed for the discrimination of reliable from non-reliable witnesses;
• Content analysis models of testimonies and of nonverbal behaviour are generally valid from laypeople’s point of view if and when they are provided with the criteria;
• Without guidelines, laypeople rest their judgement-making on several criteria such as content criteria, scientific criteria, legal criteria, and non-verbal behaviour. Furthermore, most of the criteria they use are in line with the predictions of objective models. However, they perform the task with other valueless unempirical criteria, referred to as subjective indicators.

Consequently, it is possible to train subjects to use objective indicators of content analysis and nonverbal behaviour in judgement-making. Under the assumption that it is possible to correct informal judgement-making, judges, jurors and other legal decision makers may be trained to mitigate their metacognitive deficits which have been identified as the main source of informal reasoning (Perkins, 1989). That is, once they are made aware of the subjective indicators they have used in judgement-making, it is expected they employ self correction towards objective indicators.

REFERENCES


LAYPEOPLE’S CRITERIA FOR THE DISCRIMINATION OF RELIABLE FROM NON-RELIABLE EYEWITNESSES


The influence of field dependency on eyewitness accuracy in free and cued recall

D. EMMETT, BRIAN R. CLIFFORD & P. GWYER

INTRODUCTION

A number of studies, both within the eyewitness paradigm and in other fields, have demonstrated the beneficial effect on recall of reinstating, at retrieval, the environmental context within which an event occurs, (e.g., Clifford & Gwyer, 1999; Geiselman, 1988; Gwyer 1997; Smith, 1979, 1985, 1988, 1994). Closer examination of the context reinstatement literature, however, indicates that this effect varies widely. Thomson and Davies (1988) in commenting on the variable effects found draw attention to the many ‘process’ differences between studies, in particular, the varying stimuli material used (e.g., word lists, staged events, video sequences), the different memorial tasks studied (e.g., tests of free recall, cued recall, and recognition), and the differing ways in which context is reinstated (e.g., mental, physical, and multiple reinstatements).

However, even when full account is taken of these process differences much variation remains and additional causes must be sought. If process differences can not provide a complete explanation for the variation seen in context reinstatement effects then one is drawn to the conclusion that a partial explanation may lie in the individual differences between the participants involved.

McSpadden, Schooler, and Loftus, (1988) in discussing the variation found in a series of their own context reinstatement experiments, having made due allowance for process differences, offer the suggestion that some participants may be more susceptible to context reinstatement than others. Smith & Rothkopf (1984) go further and suggest that one aspect of individual difference, which may be capable of explaining this inconsistency, is that of ‘Cognitive Style’. In particular
Witnesses

they single out Field Dependence/Independence (FDI) as having the potential to determine an individual’s susceptibility to context reinstatement. They go on to speculate that only Field Dependent individuals will benefit from context reinstatement and that unless the FDI construct is accounted for in the study of memorial context effects a complete understanding of such effects is unlikely.

FDI is a construct in which individuals are positioned along a continuum running from extreme Field Dependence (FD) at one pole to extreme Field Independence (FI) at the other (Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). FDI describes two contrasting ways of processing information. Individuals located towards the FD end of the continuum have difficulty in separating incoming information from its contextual surroundings, are more likely to be influenced by external cues, and to be non-selective in their information uptake. FI individuals have no difficulty in separating the most essential information from its context, are more likely to be influenced by internal than external cues, and to be selective in their information uptake (Ford, 1995; Hampson, 1988; Riding & Cheema, 1991; Tinajero & Paramo, 1998; Witkin, et al 1962).

Emmett, Clifford, & Gwyer (in press) established in two studies a connection between the susceptibility of eyewitnesses to the reinstatement of encoding context at retrieval and their positioning on the FDI continuum. This connection was, however, only apparent in free recall tasks where FD witnesses retrieved significantly more correct information regarding apparently genuine staged events under conditions of context reinstatement than when no such reinstatement was carried out. FI witnesses showed no such increase. Emmett et al. (in press); Eagle, Goldberger, & Breitman, (1969); Goodenough, (1976); Hosch, (1994); Smith, & Rothkopf, (1984); Witkin et al (1962) postulate that the enhanced susceptibility of FD witnesses to context reinstatement arises through their closer engagement with the environmental context surrounding the event focus during encoding within memory. They argue that this event focus becomes firmly embedded within its context during encoding, and that FDs are therefore able to make more use of the cueing provided by context reinstatement, than do FIs, in locating and successfully recalling events from episodic memory. FIs engage more closely with the focus of the event, and pay less attention to its environmental context thus encoding it at a low level of importance. If this is so, reinstating context at retrieval will, to FIs, be of limited value. Emmett et al. (in press) found no context reinstatement effects for either FD or FI witnesses in cued recall tasks and suggested that the ‘Outshining Hypothesis’ (Smith, 1988) may provide a possible explanation for this. Expressed briefly, this hypothesis proposes that the memory cueing provided by the questions in a cued question task ‘outshines’ or overpowers and makes unusable the weaker cueing provided by the reinstatement of context, in the same way that the presence of a bright light makes it more difficult to see a much weaker one in the same visual field. However, in both studies FI witnesses recalled more correct information under cued recall than did FD witnesses (p<0.05 in study one, p=0.101 in study two). In their discussion of this Emmett et al. (in press) draw attention to the high degree
of organisation imposed upon incoming information by individuals positioned close to the FI pole of the FDI continuum, and the important role potentially played by such organisation in answering detailed cued questions. Emmett et al. (in press) concentrated exclusively on items of correct information recalled and did not take account of any erroneous information produced.

Geiselman, Fisher, MacKinnon, & Holland (1985) examined the question as to whether the Cognitive Interview (Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissian, & Prosk, 1984), which uses as one of its most important strategies the mental reinstatement of encoding context, leads to an increased recall of erroneous information along with an increase in correct information. In their study, using films of simulated crimes and a forty-eight hour delay, combined free and cued recall results were compared between witnesses who were interviewed using the Cognitive Interview technique and those who were received a standard interview. Geiselman et al. (1985) found a significant increase in correct information recalled with no significant increase in erroneous information. Wilkinson (1988) in a study involving children recalling a school trip taken one day before similarly found that the reinstatement of the physical environmental context of the trip led to an increase in correct recall with no concomitant increase in errors.

The present study examines the question as to whether the differential susceptibility to context reinstatement between FD and FI witnesses demonstrated in Emmett et al. (in press) with regard to items of correct information produced in free recall, and the enhanced facility of FI witnesses in cued recall has implications for the amount of erroneous information produced. Drawing upon the results of Emmett et al (in press) and the work of Geiselman et al. (1985) and Wilkinson (1988), it seems reasonable to hypothesise that the reinstatement of context will lead to an increase in correct information for FD witness in free recall with no concomitant increase in erroneous information, and that in cued recall FI witnesses will produce more correct information and less erroneous information than FD witnesses.

2 | METHOD

2.1 | DESIGN

The experiment involved a 2 x 2 factorial design. The first factor was that of context, with two levels, ‘context reinstatement’ (CR) and ‘no context reinstatement’ (NCR). The second factor that of Field Dependency, with two levels, Field Dependent (FD) and Field Independent (FI). The FDI positioning of participants was assessed using the Group Embedded Figures Test (GEFT; Witkin et al. 1971). The median GEFT score being used to divide participants into two groups. Those whose score fell below the median were classified as FD, whilst those whose score fell either on or above the median were classified as FI (Smith & Rothkopf, 1984).
Participants were exposed to a live, apparently genuine, intrusive but non-threatening interruption to a lecture. One week later half of the participants were tested using free and cued recall under conditions of physical and mental context reinstatement (CR), the other half without any such reinstatement (NCR). In order to reinstate the physical environmental context of the interruption the CR group were tested in the same room in which encoding had occurred, and were instructed to use as an additional memory aid the contextual reinstatement mnemonic of the cognitive interview (Geiselman, 1988). The NCR group was tested in a different room from that in which the encoding phase had taken place, and without any instruction to use the context reinstatement mnemonic. They were thus not provided with any degree of context reinstatement as an aid to recalling the to-be-remembered event.

The two dependent variables of interest in the experiment were ‘free recall’ and ‘cued recall’. All participants provided data for both of these measures.

2.2 | PARTICIPANTS

Fifty-two sixth form college students, divided into two equal sized groups volunteered to take part in both the encoding and recall phases of this experiment. The combined participant group consisted of 11 males and 41 females, with a mean age of 16.99 years.

2.3 | PROCEDURE

During the course of the same afternoon both of the participant groups attended a lecture theatre with the ostensive task of completing the GEFT. The lecture theatre was equipped with tiered seating providing a clear and unobstructed view of the front of the theatre. The participants were instructed as to the requirements of the GEFT and the test completed. The test booklets were collected by the experimenter who then explained that the results of the test would be reported back to the participants one week later. As he was doing this he was interrupted by the entry of a male confederate who was not known to any of the participants. The confederate claimed to have left a number of documents in the lecture theatre and then proceeded to carry out a search of all parts of the room. He also asked the participants to check the area around them for the missing documents. The interruption was noisy and intrusive and lasted on both occasions for approximately two minutes and twenty seconds. The confederate then thanked everyone for their help and left the room. Participants were asked to gather in a designated room one week later for the results of their GEFT. To aid accuracy during the subsequent scoring of recall questionnaires, a colour photograph of the confederate was taken shortly before he entered the lecture theatre and he carried a concealed tape recorder throughout.
The retrieval phase of this experiment took place one week later. The true nature of the lecture interruption was revealed to both groups and their memories of it ascertained through the use of a two-section questionnaire. The first section of the questionnaire required participants to provide a free recall account of the interruption, a cued recall section, which posed a number of detailed questions about the interruption and provided space for an answer, followed this. In the cued section of the questionnaire it was made clear to participants that they could opt to leave any of the questions unanswered if they so wished.

2.4 | SCORING

Free recall was scored by allocating one point for each correct item of information. ‘Item’ was taken as meaning that which would have been sufficient to correctly answer one of the questions in the cued recall section, or a similar valid question that could have been included in that section. Cued recall was scored by allocating one point for each question or part of a question that was answered correctly. The number of items of erroneous information offered in both free and cued recall was also counted. In both free and cued recall the number of correct items recalled, together with the number of errors, was used to calculate a figure representing correct information produced as a proportion of the total information produced.

3 | RESULTS

<table>
<thead>
<tr>
<th>Free Recall</th>
<th>NCR</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct items</td>
<td>10.77 (2.95)</td>
<td>15.23 (4.51)</td>
</tr>
<tr>
<td>Incorrect items</td>
<td>2.85 (1.46)</td>
<td>1.38 (1.12)</td>
</tr>
<tr>
<td>Proportion correct</td>
<td>.792 (.101)</td>
<td>.914 (.07)</td>
</tr>
</tbody>
</table>

FD

| Correct items | 13.23 (3.14) | 15.54 (3.67) |
| Incorrect items | 2.54 (1.56) | 3.0 (1.47) |
| Proportion correct | .842 (.08) | .838 (.07) |

FI

Standard Deviations in parenthesis

Analysis of the three Free recall measures by 2 x 2 multivariate ANOVA indicated the following.
3.1 | CORRECT ITEMS RECALLED

A significant main effect for context condition, \( F(1, 48) = 11.38, p < 0.05 \). A non-significant main effect for FDI, \( F(1, 48) = 1.904, p > 0.05 \). A non-significant interaction between context and FDI, \( F(1, 48) = 1.152, p > 0.05 \). Significantly more correct items of information were recalled in the CR condition (mean \( 15.38 \)) compared to the NCR condition (mean \( 12.00 \)). Number of correct items recalled was not significantly affected by FDI or by context and FDI in combination.

Post Hoc analysis of the difference in numbers of correct items recalled between NCR and CR conditions by the FDI positioning of witnesses, using ‘t’ tests, indicated that the difference between CR and NCR for FI witnesses was non significant (\( t(24) = 1.724, p > 0.05 \)), and for FD witnesses was significant (\( t(24) = 2.985, p < 0.01 \)). It is the significant increase in correct items recalled by FD witnesses that powers the significant increase seen overall.

3.2 | FREE RECALL ERRORS

A non-significant main effect for context condition, \( F(1, 48) = 1.625, p > 0.05 \). A non-significant main effect for FDI, \( F(1, 48) = 2.779, p > 0.05 \). A significant interaction between context condition and FDI, \( F(1, 48) = 6.01, p < 0.05 \). The number of errors made in free recall was not significantly affected by either context or FDI but was significantly affected by an interaction of the two conditions with the number of errors made by FD witnesses falling with context reinstatement (NCR mean \( 2.85 \), CR mean \( 1.38 \)) and the number made by FI witnesses rising (NCR mean \( 2.45 \), CR mean \( 3.00 \)).

Figure 1: Free recall errors by context condition and FDI.
3.3 | PROPORTION OF RECALLED INFORMATION CORRECT

A significant main effect for context condition, \( (F (1, 48) = 6.713, p < 0.05) \). A non-significant main effect for FDI, \( (F (1, 48) = 0.309, p > 0.05) \). A significant interaction between context condition and FDI, \( (F (1,48) = 7.539, p < 0.01) \). A significantly greater proportion correct figure is seen in the CR condition (mean 0.876) compared to the NCR condition (mean 0.8166). The proportion correct was not significantly affected by FDI but was significantly affected by an interaction of the two conditions with the figure for FD witnesses rising with context reinstatement (NCR mean 0.7915, CR mean 0.8138) and for FI witnesses falling (NCR mean 0.8417, CR mean, 0.8382).

Figure 2: Correct proportion by context condition and FDI.

4 | CUED RECALL

Analysis of the three Cued recall measures by 2 x 2 multivariate ANOVA indicated the following.

4.1 | CORRECT ITEMS RECALLED

A non-significant main effect for context condition, \( (F (1, 48) = 1.751, p > 0.05) \). A non-significant main effect for FDI, \( (F (1, 48) = 3.835, p > 0.05) \). A non-significant interaction between context and FDI, \( (F (1, 48) = 0.339, p > 0.05) \). The num-
Witnesses

The number of correct items recalled was not significantly affected by either Context or FDI on their own or in combination. However, the difference between the number of correct items recalled by FD witnesses (mean 18.15) and by FI witnesses (mean 21.0) is close to significance (p = 0.056).

4.2 | CUED RECALL ERRORS

A non-significant main effect for context condition, (F (1, 48) = 1.221, p > 0.05). A non-significant main effect for FDI, (F (1, 48) = 2.573, p > 0.05). A non-significant interaction between context condition and FDI, (F (1, 48) = 0.103, p > 0.05). The number of errors made in cued recall was not significantly affected by either context or FDI or by these two variables in combination.

4.3 | PROPORTION OF RECALLED INFORMATION CORRECT

A non-significant main effect for context condition, (F (1, 48) = 0.009, p > 0.05). A significant main effect for FDI, (F (1, 48) = 5.113, p < 0.05). A non-significant interaction between context condition and FDI, (F (1, 48) = 0.345, p > 0.05). A significantly greater proportion correct figure is seen for FI witnesses (mean 0.6987) compared to FD witnesses (mean 0.6309). The proportion correct was not significantly affected by context condition or by an interaction of the two conditions.

4.4 | SUMMARY OF RESULTS

This study closely replicates the findings of Emmett at al. (in press). The significant increase in correct information in free recall under context reinstatement is seen only in FD witnesses. CR produces no significant increase for FI witnesses.
Under CR the number of errors made by FD witnesses falls whilst for FI witnesses it raises. This leads to a significantly higher proportion correct figure for FD witnesses under CR than for FI witnesses. In cued recall no significant context reinstatement effect is seen for either FD or FI witnesses, however, FI witnesses overall produce a near significantly greater amount of correct information, fewer errors, and a significantly higher proportion correct figure, as compared with FD witnesses.

5. DISCUSSION

The study has shown that, as found in other studies (Emmett et al. in press; Flexer & Tulving 1978; Groeger 1997; Memon & Bruce, 1985; Sanders, 1984), the reinstating of encoding context, at the time of retrieval, significantly increases the amount of correct free recall, but that the same phenomena is not observed for retrievals produced under cued recall. It has also replicated the findings of Geiselman et al. (1985) and Wilkinson (1988) in that the reinstatement of encoding context has not led to an increase in erroneous recall being produced. Indeed for FD witnesses the opposite is true, the hypothesis posed at the commencement of this study that FD witnesses would produce more correct and less erroneous free recall information under context reinstatement has been supported. Also supported is the hypothesis that FI witnesses would produce more correct and less erroneous information in cued recall than would FD witnesses.

The assertion that FD participants engage more closely with, and consequently encode more intensely, the environmental context surrounding the focus of an event (Emmett et al, (in press); Eagle, Goldberger, & Breitman, (1969); Goodenough, (1976); Hosch, (1994); Smith, & Rothkopf, (1984); Witkin et al (1962), argues that this event focus becomes firmly embedded within its context during encoding, and that FD participants are therefore able to make more use of the additional memory cueing provided by context reinstatement, than do their FI compatriots, in locating and successfully recalling events from episodic memory. The potential increase in feature overlap and the number of possible pathways to the encoded memory trace (Bower, 1967; Flexer & Tulving, 1978; Tulving, 1974; Tulving & Thomson, 1973) that context reinstatement provides, makes possible for the FD witness not only a significant enhancement of their ability to recall correct information but also of their ability to reject erroneous information. In other words, context reinstatement enables FD witnesses to produce significantly more reliable information. FI witnesses, on the other hand, engage with and encode more closely the event focus, to the detriment of the surrounding context. Context reinstatement at retrieval for the FI witness will be of limited value in recovering the details of the to be remembered event and will not therefore lead to a significant increase in correct information recall or make any impact upon the amount of erroneous information produced.

In discussing the cued recall results it is necessary to consider another aspect of the FDI style. Witkin et al (1962) describes the cognitive style of those individu-
Witnesses

als positioned at the FI end of the FDI continuum as being ‘analytical’ in nature, as distinct from the more ‘global’ style of those designated as FD. Witkin and his co-workers theorise that FIs impose a higher degree of organisation onto incoming impressions than do FD participants. He reports a memory trial in which a group of boys (no ages given) were questioned in cued recall style, regarding an event that occurred three years earlier. Boys who were designated as FI recalled the event in significantly more detail, and with more accuracy, than did those designated as FD. Witkin et al suggest that this was as a result of the analytical nature of their cognitive style, and the order that they impose upon incoming perceptions. Clearly in the cued recall condition the participant is being asked to recall specific details, in other words to analyse the stored memory and extract specific details from it. The degree of organisation that has been imposed upon it at encoding could at this stage play an important role in determining how successfully this extraction of detail is achieved. It seems reasonable to suggest that such a high degree of organisation will assist FI witnesses not only in recovering correct details but also in rejecting erroneous details. Asking a FI participant to analyse their memory for specifics would seem to be playing to their strengths. Asking FD participants, with their more global cognitive style, to analyse their memories for detail would seem, conversely, to be playing on their weakness.

The results of this study, which show that, in cued recall, FI witnesses outperform FD witnesses in both the amount of correct information they produce at retrieval, and also in the reliability of that information, would appear to fit in totally with the basic concepts of Field Dependency theory.

Koriat & Goldsmith (1996) assert that much of the memory research carried out since the days of Ebbinghaus (1895) has treated memory as though it were a ‘storehouse’ of discrete units of memory, each representing some detail of the original encoded event, and each potentially available for recall. An individual’s memory performance has largely been judged by how many of these discrete units of memory are correctly produced at recall. “Thus the most natural measure of memory is simply how many of the units of information originally presented can be recovered on a given memory test.” Koriat & Goldsmith (1996) p. 169. Indeed in Emmett et al. (in press) that is exactly how witness performance in both free and cued recall was measured. Koriat & Goldsmith (1996) go on to offer an alternative way of judging memory performance. They suggest that it is of more importance to consider the quality of the memories produced than to simply consider the quantity. Their ‘correspondence’ metaphor of memory proposes that an individual’s memorial performance should be judged on how closely the memories they produce ‘correspond’ to the event being recalled. This includes taking account of erroneous information produced as well as correct information. In other words it is the ‘purity’ or overall reliability of the recalled information that should be the measure of memorial performance and not simply the number of correct items. In the forensic situation a witness who confidently provides high levels of erroneous information along with information correctly recalled is of limited value, and may represent a potential danger to the investigation at hand. In laboratory based research the experimenter is aware of
the information presented at encoding and therefore also aware of any mistakes made by the witness. The investigator in the field situation does not have that advantage. Incorrect information provided by a witness has the potential to so damage an investigation that it can fail, or, more seriously, a miscarriage of justice can result. A witness who provides information that can be relied upon to contain few inaccuracies can be of greater value than a witness who produces more correct information but also at the same time a high level of erroneous information.

This study has demonstrated that the FDI positioning of witnesses and their manner of questioning can have a direct bearing on the ‘correspondence’ of their evidence to the event to be remembered. In free recall FD witnesses, under conditions of context reinstatement, will produce evidence that can be relied upon to contain a high proportion of correct information with few errors. In cued recall, on the other hand, it is the FI witness who will produce evidence with a high proportion of correct information and few errors.

REFERENCES


THE INFLUENCE OF FIELD DEPENDENCY ON EYEWITNESS ACCURACY IN FREE AND CUED RECALL


Emmett, Clifford, & Gwyer (in press) demonstrated the influence of Field Dependency on the susceptibility of eyewitnesses to context reinstatement (CR) manipulations in free recall. In two experiments, using staged events and one week delays, Field Dependent (FD) witnesses showed a significant increase in correct information recalled when encoding context was reinstated at recall, while Field Independent (FI) witnesses showed no such improvement. In cued recall, where no CR effect was apparent, FI witnesses recalled significantly more correct information than FD witnesses. The following experiment adds to our previous findings by examining both correct and erroneous information recalled as a function of Field Dependency. Two groups of 26 participants were exposed to a staged lecture interruption. Their memory of the event was tested seven days later through free and cued recall. One group was tested under conditions of context reinstatement and the other without any such reinstatement. In free recall, CR led to a significant increase in correct information and a significant reduction in erroneous information recalled by FD witnesses only. No significant change in either correct or erroneous information was seen in FI witness. In cued recall, where no CR effect was apparent, FI witnesses recalled more correct and less erroneous information than did FD witnesses, leading to a significant difference between FIs and FDs in correct information expressed as a proportion of total information recalled. The theoretical and applied issues raised by these findings are discussed.
I’ll never forget the sinking ferry

HOW SOCIAL INFLUENCE MAKES FALSE MEMORIES SURFACE

PÅR ANDERS GRANHAG, LEIF A. STRÖMWALL, JAMES F. BILLINGS

I | INTRODUCTION

The accuracy of recall for past events has been seriously questioned since Bartlett’s (1932) early and influential research on the consistency of an individual’s memory. More recently, research has provided evidence that there are multiple sources of error that may be introduced into memory from the moment of perception and storage, through retention, and at the time of retrieval. These errors are at the base of why we often find disagreement among eyewitness testimony for the same event; authors unknowingly plagiarizing from previously read work; married couples remembering differing details from their first date; one individual within a family recalling abuse while other family members do not.

Extensive research has shown that people can be led to believe that they have experienced an event, when in reality they have not (Loftus & Pickrell, 1995; Hyman & Billings, 1998; Hyman, Husband, & Billings, 1995; Porter, Yuille, & Lehman, 1999). In laboratory settings, individuals have been led to remember that they had a variety of traumatic childhood experiences, such as being lost in a shopping mall, being seriously injured by an animal, or having knocking over a punchbowl at a relative’s wedding (when in fact, according to parents’ or relatives’ reports, these events had not occurred). The participants that were capable of generating these memories (varying from 15% to 30% of the samples) also were able to describe the details of the events, adding more details as the number of interviews was increased (Hyman & Billings, 1998; Hyman et al., 1995).
In a novel study, Crombag, Wagenaar, and van Koppen (1996) investigated the malleability of people’s memory for public events; specifically, the extent to which individuals would describe a TV-film of a spectacular airliner crash in Amsterdam. They found that over 60% of the participants said they had witnessed the crash on television and could describe what they had seen, although no film of the crash existed. An additional, unexpected, finding was that women were more susceptible to the effect than were men. In a similar study, Ost, Vrij, Costall, and Bull (2002) investigated participants’ willingness to report having seen a non-existent film of the car crash in which Princess Diana was killed. The results agreed with those of Crombag et al. in that 45% of the participants in the pilot study, and 44% of those in the main study said that they had seen the film. In the present paper, we will extend on previous research by investigating how social influence affects people’s memories for public events.

2 SOCIAL INFLUENCE ON MEMORY

The studies by Asch (1951), showing how easily a person’s perception can be influenced by the opinions of others, have been paradigmatic for investigating social influence and conformity. Research on social influence has primarily focused on the susceptibility of individuals’ attitudes, values, beliefs and behaviors (Bless & Strack, 1998). Critically, there are relatively few studies investigating how social influence affects memory. Betz, Skowronski and Ostrom (1996) used the Asch paradigm to show that social pressure can influence an individual’s memory for a stimulus story. Betz et al. provided evidence that second-hand information, here used as social influence, significantly affected participants’ memories on a cued-recall test, and that participants were more likely to change their responses in the direction of the supplied information if their memory was poor. More recently, Roediger, Meade and Bergman (2001), using a somewhat different paradigm for studying how social influence creates false memories, convincingly showed that one person’s recall can be infected by another person’s memory errors (social contagion). Much in the same vein, Hoffman, Granhag, Kwong See and Loftus (2001) showed that social influence (i.e., having access to a confederate’s response) could undermine the accuracy of reality-monitoring decisions, as well as the realism in confidence.

In a laboratory study of the effects of co-witness information on the eyewitness reports of a simulated crime, Corey (2001) found that confederates (acting as social pressure) significantly influenced the participants’ reported details of the witnessed crime during a post-event interview. By adding phrases such as ‘the others all said his (the perpetrator’s) hat was blue’ (in fact, it was red) to the interview, the researcher was able to significantly alter the responses of the participants in the desired direction. The effect was found for the immediate memory of the event, and it was stable over the period of a week. Furthermore, Luus and Wells (1994) elegantly showed the malleability of eyewitness confidence by introducing different types of co-witness information to their witnesses after they had
made their identification of a perpetrator out of a (target absent) photo set. On a 10-point scale, the highest confidence scores were obtained for those witnesses who were informed that a co-witness had identified the same individual (8.77), and the lowest confidence score (3.57) was obtained for those told that the co-witness had argued that the perpetrator was not present in the array. Nine different types of co-witnesses’ information yielded very different confidence levels, and by introducing a social dimension to the identification situation, Luus and Wells were able to produce both witnesses highly confident and very uncertain. For effects of feedback on realism in confidence pertaining to eyewitnesses’ descriptions, see Granhag, Allwood and Strömwall (2000).

Thus, an individual’s recall for an event may be affected by for example co-witness hearsay. With the addition of the demand characteristics inherent within the research laboratory (found also during legal battles in the courtroom, criminal interrogations, and therapy sessions), there is ample evidence to support the notion that when individuals are motivated to remember and elaborate on phenomenologically limited, or even non-experienced events, they have the capacity to do so (Hyman & Billings, 1998; Hyman et al., 1995; Loftus, 1979; Loftus & Hoffman, 1989; Loftus & Pickrell, 1995; Porter et al., 1999).

3 | THE PRESENT RESEARCH

In the present paper we investigated the malleability of individuals’ memories for public events. Two studies are reported, both following a paradigm used by Crombag et al. (1996). In Study 1, we investigated the extent to which individuals report having witnessed a non-existent film of a disastrous event. In this instance, we employed the sinking of the cruise ship Estonia in 1994, a tragic accident where almost 900 persons lost their lives. In addition, we included the description of a second disastrous event, the crash of an aircraft (JAS 39 Gripen) at an air show in Stockholm in 1993. This event had an associated video known to have been widely exhibited in the media. We hypothesised that about half of the sample would claim that they had seen the non-existent film (cf. Crombag, et al. 1996; Ost et al., 2002).

For Study 2 we used a confederate as a source of social influence on memory, and our aim was to increase, as well as suppress, the participants’ willingness to create and report false memories. In line with the demonstrated effects of social influence on recognition memory (Betz et al., 1996), recall memory (Roediger et al., 2001) and reality-monitoring decisions (Hoffman et al., 2001), we predicted that positive social influence would increase the participants’ willingness to report having seen the non-existent film, whereas negative social influence would decrease their willingness to report a false memory.
4 | STUDY 1

4.1 | METHOD

Participants
One hundred and seven undergraduate students at Göteborg University and University of Skövde participated in the study. The sample consisted of 68 females and 39 males with the mean age of 26.5 years. All participants had lived in Sweden from (at least) 1992, which was important for the events used in the study.

Questionnaire (same for Study 1 and Study 2)
The participants filled out a questionnaire starting with background variables such as sex and age. Then, for half of the sample, followed a short description of the existent film (the JAS-film), questions regarding that film, then a short description of the non-existent film (the Estonia film) and questions regarding that film. The other half of the sample was first asked about the non-existent film (Estonia), and then about the existent film (JAS).

The questions were the same for both films. First, the participant answered the question “Have you seen this film?” (yes/no), and if they answered yes, they continued with marking, on a 10 cm long line, the clarity, richness of detail and vividness of the film (endpoints: not at all vivid to very vivid, etc). Furthermore, whether they answered if they could remember exactly where they were when they first saw the film in question (yes/no), and how many times they had seen it (once, 2-3 times, 4-5 times, more than 5 times). The participants who answered that they had seen the JAS-film and/or the Estonia-film were asked to mark, on a 10 cm long line, how certain they were that they had seen the film (very uncertain to very certain).

4.2 | RESULTS

Of the 107 participants responding to the questionnaire, 41 (38%) reported that they had seen the non-existent film of the Estonia sinking. Eighty-one (76%) of the 107 participants recalled seeing the JAS crash. A non-significant Pearson Chi Square ($\chi^2(1, N = 107) = 0.65, p = .42$) was found for the gender of those who reported seeing the Estonia-film, although there was a gender effect ($\chi^2(1, N = 107) = 6.58, p = .01$) for the JAS-film, with 90% of the males and 68% of the females reported seeing that film.

Because there were two events presented, one with film existent and one without, it was important to check the effects of the order of presentation. The main order effect (Estonia first or JAS first) for reporting seeing the Estonia-film was found to be significant ($\chi^2(1, N = 107) = 6.30, p = .012$), while for seeing the JAS-film it was not ($\chi^2(1, N = 107) = .157, p = .692$). Specifically, when the JAS film was pre-
sent in position one, 51% claimed to have seen the Estonia-film. In contrast, when the Estonia-film was presented in position one, only 26% claimed to have seen the Estonia film.

This finding led to further order analysis to uncover the source of the effect. If the JAS film was in position one, and it was remembered, then the Estonia film was also significantly recalled ($\chi^2 (1, N = 107) = 22.93, p < .001$). Of the 40 participants who remembered seeing JAS-film when placed in position one, 28 (70%) also claimed to have a memory for the Estonia-film. For the Estonia film being remembered in position one, the effect was non-significant for remembering the JAS film ($\chi^2 (1, N = 107) = 3.38, p = .066$). These results indicate that remembering seeing the JAS film first, strongly influenced the creation of a false memory (i.e., reporting to have seen the non-existent Estonia-film).

Memory characteristics. Dependent t-test analyses were performed on the descriptive responses for those who remembered seeing both the Estonia and the JAS film. All tests were found significant (all $p$’s < .05), with the recall for the JAS incident higher in each instance: clarity ($M = 4.0$ vs. $M = 1.5$), details ($M = 3.2$ vs. $M = 1.2$), and vividness ($M = 4.1$ vs. $M = 2.1$). Furthermore, the participants who claimed to have seen both the JAS and the Estonia-film were significantly more confident that they had seen the JAS-film ($M = 7.7$), than that they had seen the Estonia-film ($M = 6.4$).

Remember location. Of the 41 persons who reported having seen the Estonia film, 41% (17) reported that they remembered exactly where they were when they first saw the film. Of the 81 persons who reported having seen the JAS-film, only 15% (12) reported remembering the same thing. These proportions turned out to be significantly different ($z = 4.41, p < .001$).

Number of times seeing the film. The most frequently used category for how often the participants had seen the film was ‘2-3 times’, both for those reported having seen the JAS-film (49%), and those claiming to have seen the Estonia-film (51%). Of those reported having seen the Estonia-film, 12% claimed to have seen it more than 5 times; the corresponding number for the JAS-film was 19%.

5 | STUDY 2

5.1 | METHOD

Participants
Seventy-five undergraduate students at Göteborg University and Chalmers University of Technology (35 women and 40 men, mean age = 23.5 years) participated. All participants had lived in Sweden for at least eight of the last ten years, which was important for the events used in the study. Each participant was rewarded with SEK 80 (approx. USD 8). They were randomly allocated to one of
Witnesses

three conditions: positive social influence, negative social influence and control (no influence)

Procedure

For the two experimental conditions, the participants were instructed that they were to take part in the experiment at the same time as another participant (although they were not to collaborate). Each participant was also instructed that the other participant had called and said that she would be a couple of minutes late. This 'late participant' was actually a confederate (and was the same person for all conditions and sessions). The 'real' participant was seated in a room, and after a couple of minutes the other participant (the confederate) arrived. They were then both handed a questionnaire by the experimenter, who then left the room.

First both participants filled out a page with background variables like sex and age. Then a short description of the JAS-film (the existent film) followed, and questions regarding that film. Next, a short description of the Estonia-film (the non-existent film) followed, and questions regarding that film. In contrast to Study 1, the presentation order of the two films was not varied. However, the descriptions of the films and questions were the same as in Study 1. In addition from reading and answering her own questionnaire, the confederate kept an eye on the real participant’s progress. Immediately after the real participant had read the paragraph describing the Estonia film the confederate said: “Estonia – I can’t remember such a film” (negative social influence condition), or “Estonia – of course I remember that film” (positive social influence condition). Critically, the confederate did not express her statement as a way of trying to start a discussion, e.g., she did not try to catch the eye of the real participant. Instead, she mirrored a situation were she was ‘thinking out loud’. If the real participant commented back, the confederate did not respond further. In the control condition the participants were seated in a room without the company of the confederate. That is, they received neither negative nor positive social influence.

5.2 RESULTS

To reiterate, Study 1 showed that a high number of participants gave a positive response on whether they had seen the Estonia-film after first having reported memory of the JAS-film. Hence, we had to control for possible differences between the conditions in terms of ‘yes’ responses for the JAS-film (always in position one for Study 2). Importantly, we found no differences between the three conditions; (24 ‘yes for JAS’ in both experimental conditions and 21 in the control condition), $\chi^2 (2, N = 75) = 3.26$, p = .35.

As in Study 1, a high number of the participants remembered seeing the non-existent film of the Estonia disaster. In fact, across all conditions, 55% of the respondents claimed to have seen the film. In the positive social influence condition, as many as 76% claimed to remember the non-existent film. In contrast,
in the negative social influence condition, only 36% claimed to remember the non-existent film. In the control condition, 52% of the participants remembered the non-existent film. The three conditions differed significantly in terms of participants giving a positive response to whether they had seen the Estonia-film, $\chi^2(2, N = 75) = 8.18, p = .02$. The only significant pair-wise comparison was the one contrasting the two experimental conditions, $\chi^2(1, N = 50) = 8.12, p < .01$. A Chi Square-analysis for gender and recall of the Estonia film resulted in a non-significant finding, $\chi^2(1, N = 75) = 0.16, p = .198$.

Memory characteristics. Comparing the memory for the Estonia-film over the three conditions, one-way ANOVAs did not reveal any differences in terms of vividness, details and clarity (midscale positions for all means). In terms of confidence, we found a significant difference between the three conditions, $F(2, 40) = 3.62, p = .036$. A post-hoc analysis revealed that the participants in the control condition ($M = 7.95, SD = 2.14$) were significantly ($p < .05$) more confident in having seen the Estonia-film, than were the participants in the negative social influence condition ($M = 4.87, SD = 2.53$).

Remember location. Of the 41 participants who reported having seen the Estonia film, 35% (17) claimed to remember exactly where they were when they first saw the film. Of the 69 participants who reported to have seen the JAS-film, 30% (21) claimed to remember where they were when they first saw it. These proportions did not differ significantly ($z = -1.71, p > .05$).

Number of times seeing the film. The most frequently used category for how often the film was seen was ‘2-3 times’ for those claiming to have seen the Estonia-film (32%), and ‘4-5 times’ for those reported having seen the JAS-film (34%). Of those reported having seen the Estonia-film, 25% claimed having seen the film more than 5 times; the corresponding number for the JAS-film was 31%. A majority of the participants in the positive social influence condition (55%) claimed to have seen the Estonia-film 4 times or more, the corresponding figures for the negative social influence condition and the control condition were 33% and 31%, respectively.

6 | DISCUSSION

In the present paper, we investigated the malleability of people’s memory for a public event, and for this we used a paradigm developed by Crombag et al. (1996). In brief, we had two major aims. First, for Study 1, we sought to map the extent to which individuals claim to have seen a film of a disastrous event which they are told to have been broadcasted on TV, but in fact do not exist. Second, for Study 2, we sought to investigate how an individual’s inclination to claim a memory for such a non-existent film is moderated by social influence (in our study, a confederate’s reaction).
In agreement with our prediction, we found that a rather large part (38%) of the total sample claimed having seen the non-existent film of Estonia sinking. This result is almost in line with those of Crombag et al. (1996), who reported that over 60% remembered a non-existent film of an airline crash, and concurs with the results of Ost et al. (2001), who reported 44% of the participants claiming to recall seeing the film of the Princess Diana accident. In addition, we found an order effect showing that if the existent film was presented first - and it was remembered - it was significantly more likely that the non-existent film (presented in position two) would be remembered. In contrast to the study by Crombag et al. (1996), we found no gender differences in terms of claiming to have a memory of the non-existent film.

As recently summarised by Hyman and Loftus (2002), research on false memory can be interpreted in terms of a general framework of (i) plausibility judgments, (ii) image and narrative construction, and (iii) source monitoring errors. Applying this framework to the findings of Study 1, all three components for creating a false memory come into play. First, reasonably all participants remember the Estonia disaster. Hence, they know that the accident has occurred, and many may judge it plausible that the last minutes of the actual sinking of the ferry were caught on tape. The next step in the process of constructing a false memory is to construct an image of the sinking ferry. In this particular case, several pieces of information might add to this constructive process. As it was major news for weeks, most participants had probably seen some of the extensive documentary material shown on TV; for example, archive photos of the ferry, dramatic live-material on how passengers in lifeboats were rescued, emotional statements from survivors who had witnessed the sinking of the ferry, and underwater photos of the shipwrecked vessel. In addition, computer based animations of the accident appeared both in the newspapers and TV. All the above, in combination with reading the bogus information in the questionnaire (a few lines describing the content of the film) might very well have contributed to the creation of an image.

Finally, in order to actually claim to have seen a film of the sinking ferry, the image constructed must be made into a personal memory. In brief, the participants must incorrectly attribute the source of the image (i.e., the sinking of the ferry) to a film clip. That is, they must make a source monitoring error (see Johnson, Hastroudi, & Lindsay, 1993). Research show that source monitoring errors are more prone to occur as the retention interval increases (Johnson et al., 1993), and, in fact, our questionnaire was distributed six years after the Estonia disaster. In sum, the three steps to the creation of a false memory are reasonably to be characterized as an interactive process (Hyman & Loftus, 2002). For example, the more sources contributing to the creation of an image, the more difficult the source monitoring process; and the clearer the image, the more plausible the event. It should also be noted that a possible eagerness to please the experimenter (compliance) might have contributed to the results (see Crombag et al., 1996, Ost et al., 2002).
Furthermore, the results showed that the participants rated their memory of the JAS-film as being more detailed and higher in clarity and vividness, than their memory of the Estonia-film. These findings are in line with ideas found within the framework of reality monitoring (Johnson & Ray, 1981); that is, the suggestion that memories of real experiences are obtained through perceptual processes, and therefore richer in terms of for example perceptual and affective information, compared to memories for imagined events which are derived from an internal source.

Interestingly, 41% of those who claimed to have seen the Estonia-film claimed remembering exactly where they were when first seeing the film; the corresponding figure for those who had seen the JAS-film was only 15%. This result adds to the debate on how to explain so-called flashbulb memories, i.e., the detailed and vivid recollection people have of the occasion when they first heard about something very dramatic (Cohen, 1993). Our result is at odds with the explanation offered by for example Brown and Kulik (1982), that there is a special neural mechanism triggered by the dramatic event, which will cause the whole scene to be ‘printed’ in memory. In contrast, our participants reported having a better recall of an event that never had taken place (watching the Estonia film), compared to an event that actually had taken place (watching the JAS-film). Possibly, this can be explained by the participants actually having a flashbulb memory for the Estonia disaster, but which by now, many years later, is somewhat distorted. Our findings instead support Neisser’s (1982) idea that the preservation of flashbulb memories should be explained in terms of reconstruction, i.e., it is a result of frequent rehearsal, re-telling and traditional schemas that govern story telling (‘narrative conventions’). Furthermore, a majority of those who believed they had watched the Estonia-film claimed to have done so two times or more. In short, the fact that a person argues to have experienced a certain event at multiple occasions, does not guarantee that he or she really has experienced the event at all.

In Study 2, we used a confederate to introduce either positive or negative social influence. We managed to show that social influence can be used either to increase or suppress the creation of false memories. Specifically, three out of four participants who received positive social influence reported a false memory, whereas only one of three participants who received negative social influence reported having seen the non-existent film. The results obtained are in line with previous findings on social conformity and memory (Betz et al., 1996; Hoffman et al., 2001; Roediger et al., 2001). Noteworthy, the manner in which the social influence was introduced for the current study is to be characterized as unobtrusive. That is, the participants were not explicitly confronted with any influence (supporting or contradicting their own report). Instead, they heard another participant (a confederate) who very briefly was thinking out loud while filling out her own questionnaire. Still, the effect of social influence was very strong.

To explain the findings of Study 2, one has to add a social dimension to the framework for the creation of false memories suggested by Hyman and Loftus (2002).
Reasonably, the reason for the high degree of social conformity is probably due to both informational and normative influence (Brehm, Kassin & Fein, 1999). Informational influence, since the participants might have felt that their co-participant (the confederate) really had some knowledge about whether or not there actually exists a film showing the sinking of the Estonia ferry, whereas they themselves might have felt uncertain, due to for example the long retention interval and source monitoring dilemmas. In addition, it is rather common to assume that one’s memory is generally bad, and a low self-esteem in terms of memory performance obviously invites social conformity through informational influence (Bless & Strack, 1999; Walther et al., 2002). The results are possibly also to be explained in terms of normative influence, since the participants might have felt that one really should take care to remember tragic events of such dimension as the Estonia disaster (one of the worst accidents ever in the Baltic sea). However, it should be noted that such a ‘social desirability’-type of argument only applies to the result obtained for the positive social influence condition (where a majority claimed to have seen the film). Whereas the same argument work against the result obtained for the negative social influence condition where - contrary to any possible effects of social desirability - only one third of the participants claimed to have seen the film.

Interestingly, of the participants who received positive social influence - and reported having seen the Estonia-film - more than half claimed to have seen the film four times or more. Critically, the corresponding figures for the control condition and the negative social influence condition turned out to be much lower. This shows that not only can positive social influence create false memories, but also make people believe that they have experienced these (non-occurent) events multiple times. Supporting the reconstructive view of flashbulb-memories (Neisser, 1982; Neisser & Harsch, 1992), our results showed that it was more common to remember exactly where one was when watching something that never occurred, than where one was when watching a true event. Confirming the findings in Study 1, we found no gender differences in terms of claiming to have a memory of the non-existent film.

### CONCLUSIONS

In the present paper we have provided further evidence that individuals may very well report false memories for an entire public event. Reasonably, for the present context the plausibility of the event, the many potential sources adding to the image-construction process, as well as the problem of source monitoring, may have contributed to the creation of the false memories reported (Hyman & Loftus, 2002). Furthermore, by using a rather simple and unobtrusive manipulation, we showed that individuals’ memories are highly susceptible to social influence. Having both informational and normative social influence work in tandem, the participants’ willingness to report false memories was significantly increased. In short, our results reveal the power of social influence on the creation of false memories for entire public events.
REFERENCES


Witnesses


Confidence-accuracy calibration for absolute and relative judgements

NATHAN WEBER & NEIL BREWER

1 | INTRODUCTION

One problem faced by forensic psychologists is the identification of factors useful as predictors of eyewitness accuracy for both general testimony and eyewitness identifications. Eyewitness confidence is a variable often thought of as a potentially useful marker, particular in the context of eyewitness identifications. Despite its intuitive appeal, research over the past 30 years has consistently shown only a modest relationship between confidence and accuracy (e.g., Sporer, 1993; Wells & Murray, 1984). However, these findings were questioned recently by Juslin, Olsson, and Winman (1996). They argued that the point-biserial correlation, the index of the confidence-accuracy (CA) relationship used most frequently in identification studies, is not an appropriate measure of the usefulness of confidence as an indicator of accuracy. Instead they argued for the use of calibration, discussed in detail below, as an index of the CA relationship. A number of papers by Juslin, Olsson, and colleagues (e.g., Juslin et al., 1996; Olsson, & Juslin, 1999; Olsson, 2000) have reported results that indicate strong CA calibration but weak point-biserial correlation. Further, Brewer, Keast, and Rishworth (2002) demonstrated that, through manipulation of instructions to participants, it is possible to experimentally improve CA calibration. Thus, recent evidence appears to support the use of eyewitness confidence as an indicator of the accuracy of eyewitness identifications.

2 | CALIBRATION

Calibration is a measure of the extent to which subjective assessments of the likelihood of an event occurring correspond with the objective probability of that event occurring. In the context of eyewitness identification then, calibration
assesses the degree to which participants’ ratings, on a percentage or probability scale, of confidence in the accuracy of their identification decisions (i.e., the subjective assessment of the probability that the identification was correct) correspond with the accuracy of their identification decisions (i.e., the objective probability that the identification was correct). Perfect calibration is observed when the percentage of correct identifications at each confidence level, is equal to the percentage confidence of that level. For example, in a perfectly calibrated sample, of the decisions made with 0% confidence 0% would be correct, of the decisions made with 10% confidence 10% would be correct, of the 20% confident decisions 20% would be correct, and so on for every confidence level.

Calibration is typically assessed in three ways: examination of calibration curves, computation of the calibration statistic (C stat; sometimes referred to as the calibration index), and computation of the over/underconfidence statistic (O/U). Calibration curves are simply a plot of the proportion correct by confidence value for each confidence level. The C stat is a mathematical index of the degree to which the actual data deviate from perfect calibration. Specifically, the C stat is the weighted average of squared deviations from perfect calibration. The C stat can vary from 0, indicating perfect calibration, to 1, worst possible calibration. Finally, the O/U stat is an indication of the overall tendency to underestimate or overestimate the probability of a correct decision. The O/U stat is calculated as the difference between mean confidence and mean accuracy (both expressed on a probability scale), and can vary from –1, extreme underconfidence, to +1, extreme overconfidence. An O/U stat of zero indicates no over or underconfidence.

Calibration has two major advantages over the point-biserial correlation as an index of the CA relationship. First, the standard with which the point-biserial correlation compares actual performance is not appropriate. A perfect point-biserial correlation occurs when all of the incorrect decisions are made with the same level of confidence and all of the correct decisions are made with the same (but different to incorrect decisions) level of confidence. Obviously, this is extremely unlikely to occur in the real world (especially in a between-subjects context) and is, therefore, not a sensible criterion to use. Second, calibration provides information that is informative to members of the legal community. For example, the knowledge that the CA correlation is .30 does not help a police officer assess the likely accuracy of an eyewitness identification made with 70% confidence. On the other hand, the knowledge that confidence and accuracy are well calibrated would allow the officer to infer that an identification made with 70% has a 70% chance of being correct. For a complete discussion of the benefits of calibration over correlation, see Juslin et al. (1996) or Olsson (2000).

Despite its appeal as a measure of the CA relationship in eyewitness identification, calibration does have an important drawback. Calibration studies require a large number of observations to allow a stable calibration curve or C stat to be produced. For example, in an experiment contrasting the effect on CA calibration of two different manipulations and a control group, Brewer et al. (2002) tested
944 participants and still needed to collapse some confidence levels in order to produce a stable calibration curve. As a solution to this problem Weber and Brewer (in press) suggested the use of a face recognition paradigm to study CA calibration.

3 | THE FACE RECOGNITION PARADIGM

The major difference between the face recognition and eyewitness identification paradigms from a calibration point of view is the number of observations collected per participant. In typical eyewitness identification studies each participant is asked to make only a small number of identification decisions, usually one or two. In contrast, face recognition experiments may require each participant to make hundreds of recognition decisions. Therefore, the face recognition paradigm has two important advantages over typical eyewitness identification procedures for collecting calibration data. First, much smaller samples are required to achieve stable estimates of calibration. For example, Brewer et al. (2002), who tested 944 participants, only had about 300 data points in each of three experimental conditions, whereas Weber and Brewer (in press) tested 48 participants and collected a data set of over 1,000 observations in each of four experimental conditions. The second advantage of the face recognition paradigm is that, with an appropriate design, it allows collection of enough data from an individual participant to conduct within-subjects analyses of calibration, something that could not realistically be achieved in the eyewitness identification context.

Another important advantage of investigation of CA calibration in the face recognition paradigm is the potential for the development of new, or the extension of existing, theories of face recognition memory to allow the theories to make predictions about CA calibration. Extant face recognition memory theories do not address the CA relationship in a manner useful to applied researchers. Instead, these theories, developed in a signal detection framework, address receiver operating characteristics and tend to view confidence as an indicator of the stringency of a participant’s criterion for responding old (i.e., that they remember a face). As a result confidence is seen as a direct correlate of accuracy. Subsequently, existing face recognition theories do not adequately model the cognitive mechanism responsible for confidence judgements and, therefore, make no predictions about when confidence and accuracy will, or will not, be well calibrated. Recent research in the face recognition context (Busey, Tunnicliiff, Loftus, & Loftus, 2000) has, however, cast doubt on the generality of the assumption of a perfect association between confidence and accuracy. Thus, the extent to which confidence and accuracy are associated, and the factors that influence this association, will be necessary areas of empirical investigation in the face recognition domain in order to develop a model of confidence judgements in face recognition decisions. Obviously, a comprehensive theory of confidence judgements, especially CA calibration, in face recognition memory would be a tremendous advantage in making applied judgements about CA calibration and in guiding
research targeting issues of relevance to the applied face recognition memory context.

Yet another advantage of the face recognition paradigm is its power in investigating the relationships between accuracy and response latency, and confidence and response latency. It is commonly accepted that response time measures tend to have a high degree of inter-participant variability. As face recognition experiments allow researchers to collect multiple observations per condition from each participant, within-subjects analyses of response latency and its relationship with accuracy and confidence become practical. Thus, the face recognition paradigm allows detailed exploration of the role of response latency in cognitive models of face recognition decision processes.

4 | ABSOLUTE AND RELATIVE FACE RECOGNITION JUDGEMENTS

Until recently only one paper in the face recognition domain had measured the CA association using calibration. Specifically, Cutler and Penrod (1989) used the C stat as an index of CA association, although they did not examine calibration curves. However, a series of experiments by Weber and Brewer (in press) have recently compared the CA calibration for relative and absolute recognition judgements. These experiments serve as a useful example of the utility of the face recognition paradigm in the investigation of issues of applied importance.

The extant literature in the eyewitness identification domain (e.g., Cutler & Penrod, 1988; Lindsay, Lea, & Fulford, 1991; Lindsay & Wells, 1985) contains a number of demonstrations that superior accuracy, specifically less false identifications, is obtained through the use of sequential lineup procedures. This superiority is attributed to a difference in the type of judgement strategy thought to be most commonly used for each type of lineup. Researchers argue that when presented with a simultaneous lineup, witnesses tend to use a relative judgement strategy. That is, they select the line-up member that is closest in appearance to their memory of the perpetrator and place less emphasis on the absolute level of similarity between the perpetrator and the most familiar line-up member. In contrast, absolute judgement strategies are thought to predominate in sequential line-ups. Thus, each face is individually compared with the target face in memory and selected if it exceeds a threshold of familiarity or similarity with the perpetrator’s face. Some, albeit limited, support for this argument is provided by studies that have demonstrated that accurate subjects are more likely to report the use of an absolute judgement strategy (Dunning & Stern, 1994; Kneller, Memon, & Stevenage, 2001; Smith, Lindsay, & Pryke, 2000). Despite the evidence supporting the superior accuracy of judgements made from sequential over simultaneous lineups, little attention has been paid to the influence of judgement strategy on confidence, and none to the effect of judgement type on the CA relationship.
Based on the sample size used by Brewer et al. (2002), 600 participants seems a reasonable estimate of the minimum sample size necessary to conduct an experiment comparing CA calibration for absolute and relative judgement conditions (i.e., simultaneous and sequential lineups). In contrast a face recognition experiment, requiring each participant to complete 50 recognition trials (a relatively small number in the face recognition domain) would only require six participants to collect the same number of data points for each type of judgement, and would have the added power of a within-subjects design. Further, the use of multiple sets of stimuli randomly assigned to the absolute or relative conditions, the typical practice in the face recognition paradigm, allows researchers to be more confident that the observed results are not simply a function of the stimuli, but a real generalisable phenomenon. For these reasons Weber and Brewer adopted a face recognition approach in their investigation of the effect of judgement type on CA calibration.

The first of these experiments (Weber & Brewer, in press) investigated the impact of two factors, judgement type and confidence scale, on CA calibration. Judgement type had two levels, absolute and relative, and was manipulated by asking participants to make recognition judgements about single faces (absolute judgements) or pairs of faces (relative judgements). In the absolute judgement conditions participants were asked to indicate whether or not they had seen each face in the study phase, whereas the relative judgement conditions required participants to decide which of two faces they had seen in the study phase. After making a recognition judgement, participants were then required to indicate how confident they were that their decision was correct. The scale with which they were presented was varied in order to create two levels of the confidence scale factor. Specifically, two types of confidence scale were used, the full-scale ranged from 0 to 100% in 10% increments, and the half-scale from 50 to 100%, also in 10% increments. Both types of scale have been used in studies investigating calibration in eyewitness identification (e.g., Olsson, 2000). Participants completed four blocks of face recognition trials, each corresponding with one of the judgement type (confidence scale conditions.

Examination of the calibration curves (displayed in Figure 1) revealed two important results. First, no systematic CA association was evident for the lower range of the full confidence scale (i.e., 0 – 40%), whereas the upper range of the full-scale and the entire half-scale demonstrated a generally linear, positive CA relationship. Second, examination of the calibration curves for the upper confidence range (i.e., 50 – 100%) suggests some important similarities and differences between the two judgement conditions. Both the absolute and relative judgement conditions demonstrated an obvious positive relationship between confidence and proportion correct. Further, all four conditions displayed generally linear curves with similar slopes. Despite the similarity in shape and slope of the absolute and relative judgement curves, the two conditions displayed markedly different vertical locations. This indicates a difference in the degree of over/under-confidence between the two conditions. Specifically, the two relative judgement
Witnesses

conditions exhibited marked underconfidence, while the absolute judgement conditions demonstrated a closer correspondence between confidence and accuracy.

Although these results suggest a marked difference in CA calibration between the absolute and relative judgement conditions, the difference cannot necessarily be attributed to judgement type. The average proportion correct was found to differ significantly between the two judgement types. Specifically, the relative judgement conditions (half-scale: \( M = 0.84, SD = 0.10 \), & full-scale: \( M = 0.83, SD = 0.08 \)) displayed a higher mean proportion correct than the absolute judgement conditions (half-scale: \( M = 0.64, SD = 0.07 \), & full-scale: \( M = 0.65, SD = 0.09 \)). As the difficulty of the two judgement types was different, difficulty itself cannot be ruled out as the cause of the CA calibration difference between the judgement type conditions.

Consequently, Weber and Brewer (2003) undertook an investigation of the effect of difficulty on CA calibration in absolute and relative judgements, with the aim of determining the extent to which the calibration difference observed in the first experiment was due to the difference in the difficulty of the two types of judgement. By varying the presentation time of faces in the study phase (200, 500, or 1000 msec) three levels of difficulty were created for each of the judgement types. Based on the earlier results regarding the full- and half-range confidence scales (Weber & Brewer, in press) it was decided that only the half-range confidence scale would be used. The calibration curves obtained for each difficulty (judgement type condition) showed a number of features similar to those identified in the first experiment. Specifically, all of the curves demonstrated a generally positive, linear CA relationship and also similar slopes. Additionally, the absolute judgement conditions produced more overconfidence than the relative judgement conditions.

Investigation of the influence of difficulty on the CA association identified two important results. First, the results demonstrated that CA calibration was affected by task difficulty. Specifically, as the difficulty of the recognition task increased, judgements were made with more overconfidence (or less underconfidence). This pattern was demonstrated clearly for both the \( O/U \) and \( C \) statistics in both absolute (easy: \( O/U = 0.02, C = 0.006 \); moderate: \( O/U = 0.05, C = 0.009 \); & hard: \( O/U = 0.09, C = 0.019 \)) and relative (easy: \( O/U = -0.05, C = 0.009 \); moderate: \( O/U = -0.03, C = 0.006 \); & hard: \( O/U = 0.01, C = 0.004 \)) judgement conditions. Should this finding generalise to the eyewitness identification context, we would expect factors that influence the difficulty of the identification task, such as the time for which the perpetrator’s face is visible or the functional size of the lineup, to have an impact on CA calibration. Thus, to predict the extent to which confidence and accuracy will be calibrated in any given situation the difficulty of the task must be known. Therefore, it is important that a thorough understanding of the factors that influence the difficulty of eyewitness identifications be developed.
Second, the results provided some support for the conclusion that the type of judgement affects CA calibration, independent of task difficulty. The hard-relative condition was found to be easier than the moderate-absolute condition, but harder than the easy-absolute condition. Therefore, given the effect of difficulty on the O/U statistic described above, it would be predicted that if judgement type had no effect on CA calibration the hard-relative condition would display more overconfidence than the easy-absolute condition, and less overconfidence than the moderate-absolute condition. The calibration curves and O/U statistics revealed that this was not the case. Specifically, as expected, the hard-relative and easy-absolute curves both showed less overconfidence than the moderate-absolute. However, contrary to expectations the hard-relative and easy-absolute curves did not appear to differ in their degree of under/overconfidence. Thus, while Weber and Brewer’s (2003) study does not provide a clear cut answer to the effect of judgement type on CA calibration after the effect of task difficulty has been taken into account, it does suggest that should a meaningful difference exist, it would be in the form of a slight bias towards underconfidence in relative judgements (or overconfidence in absolute judgements). Given the effect of difficulty on CA calibration and the negligible difference between the judgement types observed in these data, no simple conclusion as to the superiority of either absolute or relative judgement strategies in eliciting well calibrated confidence judgements can be reached. Therefore, from an applied perspective Weber and Brewer’s (2003) results suggest that superior CA calibration would not be produced consistently using either a simultaneous or sequential lineup. As difficult tasks would favour the use of the sequential lineup and easier tasks the use of the simultaneous lineup, predictions about the difference in CA calibration between the two types of lineup can only be sensibly predicted in situations where the difficulty of the task is known. These results suggest that the choice between sequential and simultaneous lineups need not be influenced by consideration of the degree of CA calibration they elicit. Therefore, as the first empirical test of the difference in CA calibration between absolute and relative judgements, these findings provide a very important step in the development of a comprehensive understanding of simultaneous and sequential lineups.

CONCLUSION

These experiments provide a useful illustration of the three major ways in which face recognition research can advance our understanding of important applied issues. First, a focused and relatively efficient group of studies can be used to provide an answer (albeit a preliminary answer in this case) to a question of applied interest. Second, this type of research can be used to highlight important areas (e.g., the effect of task difficulty on CA calibration in eyewitness identification) of investigation for applied research to target. Third, this research can be used to bridge the gap between face recognition theory and applied research. For example, these experiments suggest that future research should be directed at developing a thorough understanding of the impact of task difficulty on CA cali-
vation with the aim of developing a model that will allow applied researchers and real world professionals to make judgements about the likely nature of CA calibration in specific situations. Such a model could play an extremely useful role in the legal system, guiding judgements about the utility of confidence as an indicator of accuracy for specific eyewitness identifications.


REFERENCES


The verbal overshadowing effect: in search of a chimera

BRIAN R. CLIFFORD

I | INTRODUCTION

A very robust effect in memory research is that of the beneficial effects of rehearsal. That is, the frequent repeating or revisiting of to-be-remembered information, either overtly or covertly, usually eventuates in better memory performance than when rehearsal has been prevented. This rehearsal effect has been found with all types of stimulus material, including remembering of faces (e.g. Chance & Goldstein, 1976; McKelvie, 1976; Read, 1979).

However, in a series of experiments Schooler and Engstler-Schooler (1990) demonstrated that identification of a target face was, in fact, negatively affected by previously providing a verbal description of that target face, an effect that has come to be known as the verbal overshadowing effect (VOE). These researchers have gone on to demonstrate the generality of the effect, with, among other stimuli, colours and wines (e.g. Fallshore & Schooler, 1995; Schooler, Ryan & Reder, 1996; Dodson, Johnson & Schooler, 1997).

In a meta-analysis of the VOE, Meissner and Brigham (2001) found a small, yet significant, negative effect of verbally describing a previously seen face, and later identifying it. It should be noted, however, that the vast majority of studies listed in this meta-analysis involved studies from one laboratory.

Other researchers have failed to demonstrate the VOE (Meissner, Brigham & Kelly, 2001; Memon & Rose, 2002; Memon & Bartlett, 2002; Tunnicliff & Clark, 1999; Yu & Geiselman, 1993). In our own laboratory we have routinely failed to find the effect (Clifford, Burke & Clifford, 2002; Clifford & Clifford, in prep; Clifford, Burke & Clifford, in prep). In addition, researchers who have found a VOE at immediate testing, have found the effect to disappear at delayed testing.
(Finger & Pezdek, 1999) while others have observed no such ‘release from VOE’ with delay (Schooler & Engster-Schooler, 1990).

Clearly then the VOE is a somewhat fragile effect that merits further investigation for both theoretical and practical reasons. Theoretically, the explanation of the VOE has undergone a number of revisions (see Schooler, Fiore & Brandimonte, 1997, for a discussion of the modality mismatch, the availability, the recoding interference, the strategy shift, and the transfer-inappropriate retrieval explanations of the VOE). Practically, the reality of the VOE is an important consideration. Should it prove to be a reliable and robust effect it has massive implications for every police force across the world. All police forces initially seek information from a witness or victim by asking them to describe their assailant in order to facilitate and begin their search for the perpetrator. If the VOE were a real effect, this practice would have to change.

The current paper, therefore, presents three studies, looking at both adults and children, that had the overall aim of demonstrating the presence or otherwise of a robust verbal overshadowing effect.

## EXPERIMENT 1

### METHOD

**Participants.** A total of 64 adults (40 females; 24 males) with a mean age of 28.5 years took part in the experiment.

**Design.** A 2x2x2 between-subjects ANOVA design was employed, where the factors were Description (describe the face of the perpetrator vs. describe the face of the Queen of England), Delay in identification (5mins. vs. 1-week), and Line-up Type (sequential vs. simultaneous). The dependent variable was identification accuracy (correct identification and incorrect identification).

**Stimulus and Test Materials.** The stimulus comprised a photograph of a female face (the target) presented full face measuring 10cm. by 10 cm., posed against a white background. The test material comprised a differently posed 10cm. by 10cm. photograph of the target plus similarly-sized photographs of foils that had been chosen by a non-experimental group of 30 participants as baring a close resemblance to the target in terms of age, appearance, hair colour and styling. The five foils and target photographs were compiled into either a simultaneous or a sequential line-up format for presentation purposes, with the target placed randomly in both line-up formats for each experimental participant.

**Procedure.** Participants were tested individually. The target photograph was presented for two minutes to each participant who was told to ‘study the photo carefully as I will be asking you some questions about it later’. At the end of the study phase, half the participants were handed a sheet of paper and were asked to write
down a description of the various features of the face they had just been studying. The features comprised hair, eyebrows, eyes, nose, mouth, chin, and face shape. The other half of the participants were given the same instructions except that they had to describe these facial features of the Queen of England, (a well known face, but not the target). Immediately after their description task, participants were presented with either a target-present simultaneous or sequential line-up and were asked to ‘choose the photograph of the person who you studied earlier’. All participants were told ‘the person you studied earlier may or may not be present’.

2.2 | RESULTS

As can be seen from Figure 1 there was no verbal overshadowing effect, which would have been indicated by a lower correct facial identification score for the verbal overshadowing condition relative to the non-verbal overshadowing condition.

From Figure 2 it can be seen that, not surprisingly, there was no suggestion of a ‘release from VOE’ when delayed correct identification was compared with immediate identification. Such a ‘release’ effect would have been indicated by a higher correct identification score at delay than at immediate testing.
Log linear analysis performed on the frequencies associated with the factors of description, delay and line-up type confirmed statistically that no verbal overshadowing effect was present, with Description appearing as neither a significant main nor interaction effect term in the final model. Follow up Chi Square tests declared the Description factor had no effect on immediate or delayed identification ($X^2 = 0.44, df = 1, p > .05$), nor on correct or incorrect identification ($X^2 = 1.00, df = 1, p > .05$).

2.3 | DISCUSSION

The results of Experiment 1 provide no evidence for the presence of a verbal overshadowing effect, despite the incorporation of a verbal description condition that has produced the effect in some, but not all, other researchers’ experiments (e.g. Schooler & Engstler-Schooler, 1990). In terms of practical implications, it could be argued that the use of a face-photograph was too artificial to draw direct implications for the criminal justice system, and that the ‘control’ condition of describing the face of the Queen of England was a poor control because it still involved the verbal description of a face (albeit not the target). Experiment 2 controlled for these two aspects.

3 | EXPERIMENT 2

3.1 | METHOD

Participants. A total of 40 adults (20 males; 20 females) participated in the experiment. Their mean age was 24.5 years. None had participated in Experiment 1.
**Design.** A $2 \times 2 \times 2$ mixed-ANOVA design was employed, where the between-subject factors were Description (describe target face vs. recall the lecture content presented), and Line-up format (target present vs. target absent). The within-subject factor was Delay in identification (1-hour delay vs. 1-week delay). Correct and incorrect identification were again the dependent variables.

**Stimulus and Test materials.** The target was presented live to the participants in a lecture theatre. He talked to the group for 5 minutes about the use of animals in Psychological experimentation and then walked among them distributing and then collecting a short questionnaire. At test, either a target-present or a target-absent simultaneous line-up was presented in a $3 \times 3$ photographic display.

**Procedure.** The target was introduced to the participants as a postgraduate student doing animal work. They were then exposed to a 5-minute talk on the pros and cons of using animals in psychological experimentation. The target was in full view of the participants for the full 5-minutes. Questionnaires were then given out by the target to each participant individually, and 2-minutes later, collected in again by the target. The scheduled lecture then began and continued for an hour. When it was finished, half the participants were asked to describe the face of the target while the other half were asked to write down the various pros and cons of animal experimentation detailed by the target. Immediately after this description/recall of content phase, half the participants engaged in a target present simultaneous line-up identification task while the other half engaged in the target-absent equivalent identification task. They were then dismissed and re-tested 1 week later after the next lecture in their timetabled sequence. At this second testing, those who had experienced a target-present line-up initially now experienced a target-absent line-up while those who had experienced a target-absent line-up initially now underwent a target-present line-up. In all cases participants were informed `the target may or may not be present' immediately before the line-ups were exposed.

### 3.2 RESULTS

Once again the VOE failed to appear, as can be seen in Figure 3. The frequency of correct identification is greater under the target-description than the non-target-description control condition. This is opposite to what the VOE would predict. Again, there is no suggestion of a ‘release’ effect as the frequency of correct identifications decreases with delay rather than increases, as would be predicted by the releases from VOE phenomenon.

A Log linear analysis on the factors of description, line-up type (simultaneous or sequential) and line-up format (target-present or target-absent) declared that the final model had generating class line-up format (L.R. Chi$^2$ Change $= 3.99$, df $= 1$, $p < .05$). Thus, once again, description condition was not a significant contributing factor to the observed frequencies.
DISCUSSION

Again the VOE has failed to materialise. In this study the target was live rather than mediated via a photograph, and the control condition did not involve a description of any face. Despite this there was no negative effect of verbally describing the target face on correct identification, and hence no release from the ‘detrimental’ effects of having rehearsed a face that would later be identified from a line-up.

Having failed to find a VOE in adults in two studies, the third study sought the effect in children. Children are an interesting and important class to study. They are interesting because if the VOE is predicated upon visual expertise but verbal inexpertise (Schooler et al., 1997) then it is generally accepted that children have less expertise than adults in face processing (Sonneville, Verschoor, Njiokiktjien, Veld, Toorenaar & Vranken, 2002) and tend to use featural processing rather than the more adult-like configural processing (Carey & Diamond, 1994). In terms of importance, children are an important area of study because they are appearing more and more frequently as witnesses and victims in the criminal justice system (Clifford, 2002), thus it is important to know if initial questioning about a perpetrator’s appearance decreases their later identification ability.
EXPERIMENT 3

4.1 | METHOD

Participants. A total of 120 children, comprising 60 males and 60 females took part in the experiment. Of the 120 children, 60 were aged 5/6-years of age (mean age = 5.9 years) and 60 were aged 9/10 years of age (mean age = 9.10).

Design. A 2x2x2 between-subjects ANOVA design was employed where the factors were Age (5-6 year olds vs. 9-10 year olds), Description (describe face of target vs. recall content of story), and Delay in identification (5 minutes vs. 1-week). As before, the dependent variable was correct and incorrect identification.

Stimulus and test materials. The target was presented on a video for 3 minutes telling an age-appropriate story. The target was a male 23 year old who talked to camera for the duration of the video. Test materials consisted of face-photographs of the target and five foils that matched the target in age, appearance, hair-style and dress (casual), as decided by 20 non-experimental participants. These target and foil photographs were presented as a target-present simultaneous line-up.

Procedure. Participants viewed the video in groups of five. They were instructed to ‘watch and listen to the video as you will be asked questions on it later’. After viewing the video, half the participants were asked to describe the face of the person talking on the video with the help of a checklist of features as in Experiment 1. The other half were asked to check off the correct answers to questions posed about the content of the story the man on the video had been telling. Having completed this phase, half the participants in each description condition (face/story content) proceeded to attempt an identification of the person in the video. The other half were dismissed and tested one week later. In all cases, participants were tested individually and told ‘the person you saw in the video may or may not be present’.

4.2 | RESULTS

For the first time a VOE can be seen. As Figure 4 indicates those who described the target’s face performed more poorly than those who recalled the story. In addition, there was a ‘release from VOE’ as can be seen from the comparison of immediate and delayed correct identification for face description. The opposite effect occurs for those who recalled the story – their performance decreased with lapsed time.
Figure 4: Frequency of correct identification as a function of description and delay conditions

Figure 5 indicates that the presence of a VOE and its release occurred for both age groups but is more marked for the younger (5/6) age group.

Figure 5: Frequency of correct identification as a function of description delay and age conditions
Unfortunately, when these data were subjected to a Log linear analysis the numerical differences were found to be unreliable. The analysis performed on the frequencies associated with the factors of age, description and delay declared that the best model had no factors and that a saturated model only fitted the data (L.R. Chi square = 1.770, df = 7, \( p > .05 \)). Follow up Chi Square tests indicated that while face identification was less in the face description condition this difference was not significant \( (X^2 = .529, \text{df} = 1, p > .05) \). This non-reliability of numerical differences held for both 5/6-year-olds \( (X^2 = .529, \text{df} = 1, p > .05) \) and 9/10-year-olds \( (X^2 = .048, \text{df} = 1, p > .05) \). In terms of the ‘release from VOE’, again specific follow up Chi Squares indicated that the numerical increase in accuracy at delay compared to immediate testing was unreliable \( (X^2 = .857, \text{df} = 1, p > .05) \) overall, and for both 5/6-year-olds \( (X^2 = .429, \text{df} = 1, p > .05) \) and 9/10-year-olds \( (X^2 = .200, \text{df} = 1, p > .05) \).

### 4.3 | DISCUSSION

For the first time a VOE and its release has been detected, at least numerically. However, statistical analysis revealed that this pattern was not reliable. On theoretical grounds it may have been predicted that because children do not rehearse automatically (Kail, 1990) the imposition of a rehearsal condition would disrupt their normal memory processes eventuating in poorer identification. Such seems to be the case. On the other hand, if children process faces featurally rather than configurally, then the description task, which required featural processing, should have helped the children to encode, store and retrieve the face. Such seems not to be the case. Theoretical speculation, however, must be constrained due to the absence of a statistically reliable confirmation of the suggestive data patterns.

### 5 | GENERAL DISCUSSION

The background to this presentation was uncertainty over the reality or otherwise of the verbal overshadowing effect (VOE) whereby the giving of a description of a target face is held to cause decrement in the ability to later identify that face (see Schooler & Engstler-Schooler, 1990). The literature surrounding the effect is conflicting: some find the effect and its endurance overtime; others observe the effect immediately but not eventually, while yet others do not find either VOE or its release under any conditions.

The present studies fall within the last camp. In none of the three experiments was the VOE statistically reliable. Despite using photographs, video mediated and live targets, the effect failed to appear reliably. Despite using control conditions that either involved describing a non-target face or no description of faces at all, the effect failed to materialise. In terms of a release from the VOE, logically there can be no release if the VOE does not appear initially. Both between- and within-
subject delay conditions were employed but the VOE and its release stubbornly failed to materialise.

Only when we used children as participants did the pattern of results predicated by the VOE appear, but statistical analysis indicated the pattern was unreliable.

Thus while only the third experiment produced any suggestion of a VOE, statistical analysis indicated no experiment revealed the effect reliably. Why should this be? It has been pointed out above that the VOE explanation has undergone several transformations. When the same data is seen to require several accounts one begins to lose faith in the data. A problem with several of the papers that demonstrate the effect is that it is unclear just what facial description and control tasks actually involved. This issue was looked at by Meissner, Brigham and Kelly (2001) who found that certain types of facial description tasks produced a VOE while others didn’t. In the current paper the description tasks were made explicit and no VOE appeared.

With our adult data (Experiment 1 & 2) a rehearsal-based account best explains the numerical pattern of data. If participants are asked to rehearse (describe) a target face, they will perform better than participants not asked to rehearse the target’s face. The lack of statistical difference between the face rehearsal and non-face rehearsal conditions could be explained by the rehearsal being non-optimal for adults (featural rather than configural), and visual memory being stronger than verbal memory. Thus, dual coding (Paivio, 1971) only improves adult memory for a visually encoded face marginally. The children’s data (Experiment 3) can be accounted for by the same mechanism, but in their case the imposition of a rehearsal scheme which they are known not to employ causes problems which manifest themselves in numerically poorer correct identification. In this case dual encoding interferes with the children’s more powerful visual coding of the seen face.

Irrespective of the mechanism involved in the VOE, the current three experiments add to the literature that doubts such an effect exists. The paper therefore talks to the important practical issue of whether or not psychologists should be telling police officers that the techniques they use to get their investigations off the ground are flawed and should be stopped. The import of the present paper is that there is no evidence for a verbal overshadowing effect and therefore police procedures are not flawed, and can continue.

REFERENCES


Witneses


PART 3
Victims and Offenders
Family violence and animal abuse by juveniles

THE INTRAGENERATIONAL CYCLE OF VIOLENCE

Anna C. Baldry, PhD

Children who are abused suffer short and long term consequences. Abuse can be physical, sexual or psychological. Negative consequences of abuse include psychological maladjustment, the development of aggressive behaviour or withdrawal, poor school performance, depression, anxiety, psychosomatic symptoms, and even suicidal attempt. Abuse can be directly inflicted or witnessed as in the cases of exposure to domestic violence.

I | HIGH RISK GROUP

Kolbo et al. (1996) and subsequently Edleson (1999), reviewed studies on how domestic violence affect children exposed to it. They found that boys exposed to domestic violence are more likely to become aggressive and exhibit antisocial behaviour (the so called ‘externalised behaviours’) than non-exposed ones. Cruelty against animals is often included as a component of ‘acting out’ problems (Ascione, 2000) and an index of conduct disorder (together with fire setting and vandalism) connected with direct abuse. Exposed girls, instead, are more likely to become submissive, exhibiting the so called ‘internalised’ behaviours, such as depression. Factors found to moderate the negative impact of exposure to violence include whether the child was also abused, child gender, age and length of exposure to violence (Edleson, 1999). Children exposed to violence are a ‘high risk group’ for the development of psychopathology.

According to the American Psychological Association (APA, 1996: 57): “Children who witness parental violence have reactions similar to those of children who are
Victims and offenders

direct victims of abuse”. “A child’s exposure to the father abusing the mother is the strongest risk factor for transmitting violent behavior from one generation to the next” (APA, 1996: 53). According to the review conducted by Widom (1989), up to 70% of violent adults have had a history of child abuse (either as direct victims or witnesses of violence). Retrospective studies with samples of adult offenders have indicated that around 30% of them had been living in violent families; prospective studies indicate that around 15% of children who witness violence or are directly abused become delinquents. Violence produces violence: children living with domestic violence or those who were been abused are at a higher risk of developing aggressive and violent behavior than those who have not (Widom, 2000).

Exposure to violence, however, could be not only in the family, but also in the community, on television, exposure to violence by peers at school and in the streets; violence perpetrated against people but also against animals. Ascione (1993, 1998) suggested that witnessing animal cruelty can increase the child’s risk of developing future externalizing problems including violence against people and animals. According to social learning theory, violence is learned as any other form of behavior from very early in childhood (Reitzel-Jaffe and Wolfe, 2001).

Ascione (1993, 1998, 2001) was among the first researchers who investigated the link between domestic violence and exposure to cruelty against animals. He surveyed a sample of 38 women seeking help in a shelter in northern Utah for battered women using his Batter Partner Shelter Survey (BPPS) – Pet Maltreatment Assessment (Ascione, Weber & Wood, 1997). Women were asked about the prevalence of violence against their pets by their violent partners and whether also the children committed any abuse or witnessed not only violence in the family but also violence against their beloved pet companion. Results indicated that 71% of the women who had a pet said that their partner abused the animal or threatened to do so, and abuse of animals by children was reported in 32% of cases of all battered women who had one or more child. This study was limited in size and representativeness of the sample because respondents were recruited from only one shelter. In a later study, Ascione, (2000) compared animal abuse by partners among women in shelters and in the community and found that battered women were about three times more likely to report partners’ animal abuse compared to community women. Flynn (2000a) found slightly lower figures of animal abuse by partners, showing that 46.5% of the 44 women in a shelter in South Carolina who reported having pets, had batterers who threatened to harm or harmed the pet. Less evidence was reported of children harming their pets in this study.

Cruelty against animals is perpetrated by violent abusers as a way to threaten their partners and exercise power by inflicting pain to weaker creatures; children for their part, are cruel to animals as a consequence of exposure to traumatic
events. They might also have witnessed others doing so (in the family or by peers) and learned it (Ascione, 1998). Children are may be more likely to be cruel to animals when they are with other peers than when alone. ‘Peer reinforcement for “showing off” or “daring” may result in collective cruel acts to an animal that would not occur if the child were alone’ (Boat, 1995: 230). Watching friends being cruel to animals and abusing them accordingly makes them feel more powerful and this reflects their desire to control and inflict pain. Animal abuse, in this respect, is an indicator of maladjustment; by using violence against animals youngsters can regain a sense of control and power to compensate for their own deprivation from affection and care or else in cases in which they have been target of direct abuse.

There is evidence of a strong relationship between corporal punishment against the child and children’s cruelty to animals (Flynn, 2000b). In the study conducted by Flynn (1999) with a sample of 267 US college undergraduates, it emerged that boys committing animal cruelty were more likely to have suffered corporal punishment at the hands of their fathers compared to non animal abusers.

There are several possible explanations provided of why adult abusers are cruel against animals (Kellert and Felthous (1985):

- To control an animal (i.e., animal abuse as discipline or “training”);
- To retaliate against an animal;
- To satisfy a prejudice against a species or breed (e.g., hatred of cats);
- To express aggression through an animal (i.e., training an animal to attack, using inflicted pain to create a “mean” dog);
- To enhance one’s own aggressiveness (e.g., using an animal victim for target practice);
- To shock people for amusement;
- To retaliate against other people (by hurting their pets or abusing animals in their presence);
- To displace hostility from a person to an animal (i.e., attacking a vulnerable animal when assaulting the real human target is judged too risky);
- To experience non-specific sadism (i.e., enjoying the suffering experienced by the animal victim, in and of itself).

Regarding explanations of why children and adolescents abuse animals (Ascione, Thompson, and Black, 1997), several causes can be identified, often provided as excuses provided by young abusers themselves:

- Curiosity or exploration (i.e., the animal is injured or killed in the process of being examined, usually by a young or developmentally delayed child);
- Peer pressure (e.g., peers may encourage animal abuse or require it as part of an initiation rite);
- Mood enhancement (e.g., animal abuse is used to relieve boredom or depression);
Victims and offenders

• Sexual gratification (i.e., bestiality);
• Imitation (i.e., copying a parent’s or other adult’s abusive “discipline” of animals);
• Forced abuse (i.e., the child is coerced into animal abuse by a more powerful individual);
• Attachment to an animal (e.g., the child kills an animal to prevent its torture by another individual);
• Animal phobias (that cause a pre-emptive attack on a feared animal);
• Identification with the child’s abuser (e.g., a victimized child may try to re-gain a sense of power by victimizing a more vulnerable animal);
• Posttraumatic play (i.e., re-enacting violent episodes with an animal victim);
• Self-injury (i.e., using an animal to inflict injuries on the child’s own body);
• Rehearsal for interpersonal violence (i.e., “practicing” violence on stray animals or pets before engaging in violent acts against other people);
• Vehicle for emotional abuse (e.g., injuring a sibling’s pet to frighten the sibling).

Interconnection of different forms of abuse

Figure 1: Interconnection of different forms of abuse

Exposure to violence includes not only direct ‘eye’ witnessing violent events. Children can experience adult domestic violence in several other ways. Ganley & Schechter (1996) reported in Edleson (1999: 841), indicate among others: “hitting or threatening a child while in his or her mother’s arms, taking the child hostage to force the mother’s return to the home” or some forms of stalking behaviors also directed at children “who are constantly asked about their mother’s life: who she sees, where she goes, what she does” (Baldry, 2002). Exposure to animal abuse, instead, can take the form of the death or harm of a pet or other animals by the hands of a parent, a friend or other adults, or seeing the animal suffering after a cruel act perpetrated against him. The term ‘exposure’ is then preferred and used in the present article because youngsters were asked about violence between their parents regardless of whether they have actually seen such violence or not.

Exact figures on children exposed to violence are difficult to determine. According to NCH Action for Children (1994) more than half of the physical attacks against women take place in front of their children. United States estimates on family violence indicate that between 3.3 and 10 million children each year are at risk of exposure to domestic violence (Carlson, 1984; Straus & Gelles, 1990). More recent data based on the 1993 Violence Against Women Survey in Canada found that 39% of women who suffered from domestic violence during their life time reported that their children witnessed the violence (Canadian Centre for Justice Statistics, 2001). Similar findings were found by the Canadian 1999 General Social Survey (GSS), indicating that 37% of households who suffered from spousal violence in the previous five years had children who witnessed such violence. The Finnish International Survey on violence against women indicated a similar proportion: in 40% of all violence cases in partner relationships, the child/children had in some ways witnessed the violence (Heiskanen & Piispa, 1998).

These estimates are based on parents’ accounts only (primarily mothers) and refer mainly to direct eye-witnessed or heard violence, yielding to an underestimation of the real proportion of exposure with reference also to other forms covert violence, such as emotional or psychological abuse (Salcido Carter, Weithorn, & Behrman, 1999). In spite of the parents’ intention to shield children from exposure to the attacks, most children witness them (Sipe & Hall, 1996). Parents might wrongly think that their children are not aware of what is going on, either because they think they are sleeping or playing (Jaffe, Wolfe & Wilson, 1999), or else they might fear what might happen to them if they reveal such information (Dauvergne & Johnson, 2001). Mothers, in addition, tend to underestimate the impact of violence on their children and minimize or deny the amount of violence they have been exposed to, because of the so-called ‘adultification’ processes which include shifting the blame from the abuser to the children, minimizing the need for protection of the victim and attributing to children capacities they do not have (Stephens, 1999).
VICTIMS AND OFFENDERS

The same underestimation applies in cases of exposure to animal abuse when parents (mothers) are asked about the extent of violence they are aware of. The best approach to learn about children’s exposure to domestic violence is to ask directly those who might be exposed to violence. O’Brien, John, Margolin and Erel (1994: 58), in fact, found that 10% of all children in their sample reported physical aggression when both parents claimed that such violence never occurred. Children’s accounts are not a perfect criterion for assessing whether they have been exposed to violence, but they provide reliable and accurate information about what goes on between parents (Grych, Seid & Fincham, 1992), and the same could be true for abuse against animals. Offord, Boyle, and Racine (1991), in fact, surveyed a nonclinical sample of 1,232 Canadian parents/guardians and their children aged 12- to 16-year-old. They asked respondents (both parents/guardians and adolescents) to report on a number of Conduct Disorder symptoms among which cruelty against animals. Findings from this study suggest that parents and guardians may seriously underestimate cruelty to animals, with “boys self-reporting this behaviour at 3.8 times the rate of parents/guardians and girls at 7.6 times the parent/guardian rate” (reported in Ascione, 2001: 3).

When researching these issues, questions should be formulated in a sensitive way. According to O’Brien et al. (1994), ‘marital physical violence is a low base rate event [i.e. does not occur frequently] and the presence of observers often precludes its occurrence, marital physical aggression is difficult to measure through any method other than self-report’ (O’Brien et al., 1994: 46). For this reason and for ethical committee restrictions, children are rarely asked directly about any violence they are exposed to.

THE PRESENT STUDY

Participants in the study were 1396 Italian students (45.9% girls and 54.1% boys) recruited from 13 different schools recruited in the city and province of Rome: four elementary (28.5%), six middle (47.3%) and three high schools (24.1%). The age range varied from 9 years to 17 (mean age=12.1 years, SD=2.6). Students were approached in the schools in their own classes and were told that all information gathered would be used for research purposes only. Data were collected with a self-reported questionnaire measuring animal abuse according to a scale developed by Baldry (2002) measuring different types of abuse against animals by youngsters as well as from other adults including mother, father or peers. Another part of the questionnaire measures exposure to domestic violence by the father against the mother and the reverse. Youngsters were asked to indicate on a five-point scale how often their father or their mother did any of the behavior listed that ranged from name calling to hitting, grabbing or shoving.

Results reported a considerable high proportion of abuse against animals: 50.8% of all youngsters admitted committing at least one type of animal abuse (cruelty, hitting, tormenting, bothering or harming); of all these animal abuse cases,
66.5% were committed by boys, and 33.5% by girls ($\chi^2 = 86.19$, df = 1, $p<.00001$). Gender comparisons are presented for each type of animal abuse in Table 1, and they all indicate significant differences, with boys being at least twice to three times more likely to commit some kind of animal abuse compared to girls, and older boys from high school reporting more involvement than younger ones.

**Table 1.** Percentages of youngsters reporting different types of animal abuse and exposure to domestic violence between parents, overall and according to gender and school grade differences

<table>
<thead>
<tr>
<th>Animal abuse direct &amp; exposed (P.E.T. Scale)</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>All students (1536)</td>
<td>Girls (622)</td>
</tr>
<tr>
<td>Child bothering animals 34.8</td>
<td>20.6</td>
</tr>
<tr>
<td>Child tormenting animals 30.3</td>
<td>21.9</td>
</tr>
<tr>
<td>Child harming animals 19.5</td>
<td>7.4</td>
</tr>
<tr>
<td>Child cruel to animals 14.8</td>
<td>7.8</td>
</tr>
<tr>
<td>Child hitting animals 13.9</td>
<td>9.0</td>
</tr>
<tr>
<td>Father harming animals 9.0</td>
<td>6.2</td>
</tr>
<tr>
<td>Mother harming animals 5.1</td>
<td>4.2</td>
</tr>
<tr>
<td>Peer harming animals 61.7</td>
<td>57.3</td>
</tr>
<tr>
<td>Another adult harming animals 60.9</td>
<td>60.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposure to domestic violence</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Father verbally insulting mother 39.6</td>
<td>40.3</td>
<td>39.0</td>
<td>.23</td>
</tr>
<tr>
<td>Mother verbally insulting father 36.1</td>
<td>38.8</td>
<td>33.7</td>
<td>3.69</td>
</tr>
<tr>
<td>Father hitting mother 9.4</td>
<td>8.5</td>
<td>10.2</td>
<td>1.14</td>
</tr>
<tr>
<td>Mother hitting father 4.5</td>
<td>4.6</td>
<td>4.4</td>
<td>.37</td>
</tr>
<tr>
<td>Father harming mother 7.7</td>
<td>7.9</td>
<td>7.3</td>
<td>.10</td>
</tr>
<tr>
<td>Mother harming father 4.7</td>
<td>4.8</td>
<td>4.7</td>
<td>.01</td>
</tr>
<tr>
<td>Father threatening mother 7.1</td>
<td>7.3</td>
<td>6.9</td>
<td>.08</td>
</tr>
<tr>
<td>Mother threatening father 5.9</td>
<td>5.8</td>
<td>5.9</td>
<td>.01</td>
</tr>
<tr>
<td>Father throwing things to mother 5.7</td>
<td>5.8</td>
<td>5.6</td>
<td>.02</td>
</tr>
<tr>
<td>Mother throwing things to father 5.9</td>
<td>6.1</td>
<td>5.7</td>
<td>.13</td>
</tr>
</tbody>
</table>

Note: Prevalence is the percentage of children reporting whether parents, themselves or a peer ever committed any of the act listed. Percentages refer to proportion of youngsters in each category reporting different types of violence. Questions are referred to any animal.

*p < .01; **p < .001; ***p < .0001. Due to the sample size a more strict level for acceptance of significance was adopted.

Significant levels are corrected with Bonferroni Test. N in each category varies due to missing values.
To explore the independent contribution of the different types of violence a youngster is exposed to in predicting animal abuse, we performed a regression analysis. The predicting factors were: father to mother violence, mother to father violence (\(fvm\) and \(mvf\) respectively, including verbal violence), father harming animals, mother harming animals, peer harming animals, adult harming animals, age, gender and SES.

Table 2. Simultaneous Multiple Regression analysis for variables predicting animal abuse by youngsters (\(N=1310\))

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>Gender</td>
<td>-.23</td>
<td>.03</td>
</tr>
<tr>
<td>SES</td>
<td>-.01</td>
<td>.02</td>
</tr>
<tr>
<td>Parents living together</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>Father animal abuse</td>
<td>.12</td>
<td>.03</td>
</tr>
<tr>
<td>Mother animal abuse</td>
<td>.15</td>
<td>.04</td>
</tr>
<tr>
<td>Peer animal abuse</td>
<td>.10</td>
<td>.01</td>
</tr>
<tr>
<td>Adult animal abuse</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>MVF</td>
<td>.21</td>
<td>.05</td>
</tr>
<tr>
<td>FVM</td>
<td>.01</td>
<td>.04</td>
</tr>
</tbody>
</table>

Note: Fvm = father to mother violence, Mvf = mother to father violence. Fvm and mvf include also verbal violence. Gender is coded 0=boys, 1=girls; negative B is in the direction of being a boy. Full model statistics: \(R^2=.21, F(10, 1177)=31.17, p<.0001\). Differences in N’s are due to missing values.

Table 2 shows the contributing amount of variance of each predictor that was entered in the analysis simultaneously, together accounting for 21% of the total variance of animal abuse, \(F(10, 1177)=31.17, p<.0001\). Peers harming animals and being a boy were the best predictors of animal abuse \((s=.21, .21, \text{respectively})\), followed by exposure to mother violence against father and mother abuse against animals \((s=.14, .10, \text{respectively})\) and father abuse against animal and being older \((s=.10, .07, \text{respectively})\).

5 | DISCUSSION

Violence among peers is a highly reported type of violence affecting almost half of the sample interviewed. Although not all forms of abuse could be considered serious offences, still milder forms of abuse should be taken into account when considering intervention programs. The strongest predicting factor for animal
abuse is being a boy and peer animal abuse followed by exposure to other forms of abuse. Exposure to interparental violence is negatively affecting behavior meaning that these kids whose mothers are aggressive against the father or the reverse, are more at risk of becoming violent against more vulnerable creatures. The same holds in cases of exposure to the same type of violence, meaning parents abusing animals. Psychologists, psychiatrists, criminologists, sociologists, are all asked to recognize animal abuse and cruelty against animals in general as a strong (risk) indicator of maltreatment not only negatively affecting the animal, but also because it is strongly associated to abuse inflicted to children or witnessed by them. In terms of policy intervention, it is essential to gather all experiences and skills to provide intensive and effective responses to early identify high at risk cases.

REFERENCES


Victims and offenders


Rape responsibility attributions as a function of victim attractiveness and race

NORINE L. JALBERT & HEATHER SNIFFIN

1 | INTRODUCTION

According to Just World Theory (Lerner, 1970), people have a defensive attributional tendency to view the world as a just place where people “get what they deserve and deserve what they get.” Such a seemingly simplistic worldview enables us to go about our daily activities with the confidence that the world is not, after all, a random place, and we can insure our own positive outcomes and avoid negative outcomes by being good persons and/or by engaging in appropriate behaviors. This just world attributional tendency is most likely to be activated when we are confronted by an “innocent” person who suffers some kind of negative event (Lerner & Simmons, 1966), because such events threaten our belief in a just world and make us feel vulnerable.

Rape victimization is one area where just world attributions are likely to evidence themselves. Statements attributing responsibility or blame to the victim for her victimization (e.g., “she asked for it” or “what was she doing out by herself at 2:00 in the morning”) can be interpreted as just world responses. If the victim is viewed as responsible for her victimization, then the perceiver can avoid a similar negative event by not being like the victim.

Research on rape victimization and attributions of responsibility has sought to uncover the variables that might influence an observer’s tendency to make just world attributions. One underlying motive for the research focus on subjects’ attributions of responsibility to the victim is the suspicion that subjects who attribute greater responsibility or blame to the victim may, conversely, attribute less responsibility or blame to the perpetrator. One line of research focuses on varia-
Victims and offenders

bles that are relevant to the legal aspects of the rape incident, such as likelihood of victim consent, degree of victim provocation, and extent of injury or harm to the victim (Borgida & White, 1978; Kerr & Turner-Kurtz, 1977; Scroggs, 1976). A second line of research focuses on personal, “exarelegal” victim characteristics, such as respectability or social status, previous sexual experience, prior victimization, and amount of resistance (Calhoun, Selby, & Warring, 1976; Cann, Calhoun, & Selby, 1979; Jones & Aronson, 1973). Also included in this second line of research are studies investigating the role of the victim’s physical attractiveness on subjects’ attributions of responsibility (Calhoun, Selby, Cann, & Keller, 1978; Deitz, Littman, & Bentley, 1984; Gerdes, Dammann, & Heilig, 1988; Jacobson, 1981; Kanekar & Nazareth, 1988; Seligman, Brickman, & Koulack, 1977; Seligman, Paschall, & Takata, 1974; Shaw, 1971; Thornton, 1977; Thornton & Ryckman, 1983; Villemur & Hyde, 1986).

While there is widespread agreement that victim physical attractiveness does have an influence on subjects’ attributions of responsibility, there is less agreement on just what that influence is. Research by Calhoun, et al. (1978) suggests that greater responsibility is attributed to the more attractive victim, while Best and Demmin (1982) and Thornton (1977) report that victim attractiveness has no effect on responsibility attributions. The bulk of more recent research, however, (Deitz, et al. 1984; Gerdes, et al., 1988; Kanekar & Nazareth 1988; Seligman, et al., 1977; Thornton & Ryckman 1983; Villemur & Hyde 1986) seems to suggest that unattractive victims are viewed as more responsible than attractive victims and, furthermore, that the effect of victim physical attractiveness on responsibility attributions interacts significantly with other victim characteristics.

It would appear, then, that it is the unattractive rape victim who is most likely to evoke a just world response from subjects. One explanation for why subjects attribute less responsibility to the attractive victim is that subjects in this situation are confronted with two conflicting tendencies – the tendency to invoke just world attributions and the tendency to view what is beautiful as good. According to Dion, Berscheid and Walster (1974), there exists a “what is beautiful is good” stereotype that leads us generally to view physically attractive people more favorably that physically unattractive people. In a rape situation, the unattractive victim evokes only a just world response whereas the attractive victim evokes a just world response “softened” by the “what is beautiful is good” response.

One victim characteristic that seems to be missing from the research literature is victim race and the extent to which race independently or interactively with attractiveness affects male and female subjects’ attributions of responsibility. Research on cross-racial identification suggests that race is a salient variable which influences identification accuracy (Bothwell, Brigham, & Malpass, 1989; Wells, 1987; Penrod, Loftus, & Winkler, 1982), so it is only a short step to then hypothesize that race could be an equally salient variable in influencing subjects’
attributions of responsibility. The present study was designed to investigate the effect of victim race (white or black) and victim attractiveness (attractive or unattractive) on male and female subjects’ perceptions of a rape.

2 | METHOD

2.1 | SUBJECTS

Subjects were 221 predominantly white volunteers from eight sections of Introductory Psychology (103 females and 118 males).

2.2 | PROCEDURE

Each of the eight participating classes was randomly assigned to one of four conditions: white attractive victim, black attractive victim, white unattractive victim, or black unattractive victim, and the same female experimenter worked with all classes. Subjects in each class received identical test booklets consisting of a consent form, a biographical information form, a brief description of a rape incident accompanied by a photograph of the victim, and a 15-item questionnaire about their reactions to the rape incident. Victim characteristics were manipulated by using four pretested photographs of white and black, attractive and unattractive females.

The following description of a rape incident was given to all subjects with the only difference being the accompanying photograph:

The young woman in the picture above is a 20-year-old undergraduate student majoring in the Social Sciences. She is 5’4” tall and weighs 125 pounds. On February 1, 1991, she was raped in the parking lot of her apartment building. The young woman was returning to her apartment from a late night class when a man whom she did not recognize approached her. The man grabbed her, and a struggle occurred during which the assailant threatened the victim with bodily harm. The victim was forced back into her car, her clothes were stripped off, and she was forced to engage in sexual intercourse with the assailant. A passerby noticing the event, called the police who arrived and apprehended the suspected assailant two blocks from the apartment complex. Medical examinations of the victim revealed no physical injuries but confirmed that sexual intercourse had taken place. The alleged assailant was a 25-year-old white male, 5’10” tall, weighing 175 pounds. He had no prior history of sexual assault.

After reading about the rape incident, subjects were asked to complete a 15-item questionnaire about their perceptions and responses to the event. All questionnaire items were scored on an 11-point scale. The questionnaire included an item
to check the victim attractiveness manipulation and 14 items to assess subjects’ personal reactions to the victim and defendant, the degree to which subjects identified with the victim and the defendant, the level of responsibility attributed to the victim and defendant, the perceived severity of the rape consequences, the seriousness of the crime of rape, the extent to which the victim may have encouraged the rape, the degree to which the victim could have prevented the rape, the role of chance in the attack, subjects’ certainty that the correct person was arrested, subjects’ certainty of the defendant’s guilt, and subjects’ recommended sentence if the defendant were to be found guilty. Test booklets were collected when all subjects had completed the questionnaire, and the experimenter answered all subjects’ questions.

3 | RESULTS

Subjects’ responses on the 15-item questionnaire were analyzed using a 2 x 2 x 2 (race by attractiveness by subject gender) analysis of variance.

Manipulation Check. Subjects’ ratings of the victim’s physical attractiveness yielded a highly significant main effect for attractiveness \(F(1,210)=256.11; p<.0001\), and attractive victims were given higher scores \(x=7.80, \text{sd}=2.00\) than unattractive victims \(x=3.69, \text{sd}=2.07\). Significant main effects were also found for victim race and subject gender such that white victims \(x=6.40, \text{sd}=3.18\) were viewed as more attractive than black victims \(x=5.59, \text{sd}=2.52\) and female subjects rated victims as higher in attractiveness \(x=6.80, \text{sd}=2.49\) than did male subjects \(x=5.28, \text{sd}=3.02\).

Analysis of Dependent Measures. Gender differences consistent with earlier research were found. As compared to male subjects, female subjects identified more with the victim \(F(1,212)=8.20; p<.005\) and less with the defendant \(F(1,212)=13.82; p<.001\); rated the crime of rape as more serious \(F(1,213)=3.84; p<.05\) and did not feel that the victim could have prevented the attack \(F(1,212)=6.43; p<.01\). Although not significant, there was also a tendency for female subjects to view the attack as more due to chance than male subjects and to view the victim as less able to prevent the rape, less likely to have encouraged the rape, and less responsible for the rape.

The analyses of the victim attractiveness manipulation also yielded results consistent with earlier research. Subjects responded more positively to the attractive victim \(F(1,211)=6.54; p<.01\), viewed the crime as more serious for the attractive victim \(F(1,213)=8.07; p<.005\), and viewed the defendant more positively when the victim was unattractive \(F(1,211)=4.15; p<.05\).

Aside from the main effect mentioned earlier of victim race on ratings of the victim’s physical attractiveness, there were no main effects due to the race of the victim. However, even though the manipulation of victim race did not reveal any
main effects, there were several interaction effects which, taken together, sug-
est that victim race may nonetheless still be a salient variable in subjects’ percep-
tions of a rape. On the item dealing with rape prevention, there was a significant race by attractiveness effect ($F(1,212)=4.55; p<.05$) such that white attractive and black unattractive victims were viewed as more responsible. On the same item, an attractiveness by subject gender effect was found ($F(1,212)=5.16; p<.05$) indicating that male subjects, compared to female subjects, attributed about the same level of responsibility to victims regardless of attractiveness, whereas female sub-
jects attributed noticeably less responsibility in the attractive victim condition. The item on defendant responsibility yielded a race by attractiveness interaction ($F(1,210)=3.68; p<.06$) in which defendants were held more responsible when the victim was either white and attractive or black and unattractive. Two race by subject gender interactions were found. On the item dealing with recommended sentences ($F(1,205)=4.60; p<.05$), male and female subjects did not differ in their recommendations for the white victim, but female subjects were more severe than male subjects when the victim was black. Consistent with the sentencing recommendation, the item asking subjects to rate their identification with the victim yielded a race by subject gender interaction ($F(1,212)=5.13; p<.05$), and females more than males identified with the black victims, but this gender diffe-
rence did not show itself in the white victim condition.

Significant three way interactions were found on the items dealing with certain-
ty that the correct person was arrested ($F(1,212)=4.68, p<.03$), certainty that the defendant was guilty ($F(1,212)=4.49, p<.04$), recommended sentence for a guilty defendant ($F(1,205), p<.10$), and the role of chance in the rape attack ($F(1,206)=3.86, p<.05$). Subjects’ responses to each of these items revealed a consistent pattern of differential responses between male and female subjects. Male subjects attributed more chance and were more certain that the correct person was arrested and was guilty when the victim was either white and unattractive or black and attractive, yet their recommendations for sentencing were harsher when the victim was white regardless of attractiveness. Female subjects, on the other hand, attributed more chance and were more certain that the correct person was arrested and was guilty when the victim was either white and attractive or black and unattractive, yet their sentencing recommendations were harsher when the victim was black regardless of attractiveness.

4 | DISCUSSION

In general, the results of this study were consistent with earlier research investiga-
ting the influence of victim attractiveness and subject gender on attributions of responsibility. The physical attractiveness of a rape victim continues to be a salient variable for subjects, who attribute greater responsibility to unattractive victims than to attractive victims. Subject gender continues to influence subjects’ percep-
tions of and reactions to a rape incident such that female subjects identify and symp-
pathize more with the victim and less with the defendant than do male subjects.
Victims and offenders

The inclusion of victim race as a third variable failed to produce any significant main effects other than the one found on the victim attractiveness manipulation check. On that item, subjects rated the white victims as more attractive than the black victims. This effect of victim race on attractiveness ratings may be due to real differences in the attractiveness of the victim photos since only one pretested photo was used in each condition. On the other hand, the differences in attractiveness ratings of white and black victims may also reflect different standards of beauty or unconscious bias for one's own racial group.

Even though there were no main effects due to victim race, there were some interaction effects that warrant further study. When dealing with a white victim, subjects’ responses seemed to follow patterns reminiscent of earlier studies, and white unattractive victims tended to be viewed as more responsible than white attractive victims. However, a different pattern emerged when dealing with a black victim, and black unattractive victims were held less responsible than black attractive victims, especially by female subjects. In some instances the sympathies of the female subjects toward the black unattractive victim exceeded the sympathies extended to not only to the white unattractive victim but the white attractive victim as well. One explanation for this tendency might be that the just world beliefs and tendencies of the predominantly white female subjects were less likely to be elicited by the racially dissimilar victim. The greater the perceived similarity of a victim to a perceiver, the greater should be the threat to the perceiver’s just world belief when the victim is victimized. If subjects in this study focused their attention on the racial dissimilarity between the victim and themselves, there would have been less “need” to invoke just world attributions.

Another possible explanation for the greater sympathy of female subjects toward black unattractive victims (and, conversely, lesser sympathy for black attractive victims) is that racial attitudes may have consciously or unconsciously affected subjects’ perceptions. Feelings of sympathy toward an “underdog” minority member may be heightened when the victim is unattractive and lessened when the victim is attractive. Alternatively, subjects could have been reacting to the inter-racial aspect of the sexual assault in the black victim condition. In choosing to hold constant the characteristics of the alleged attacker, the researchers created a same-race assault for white victims and an inter-racial assault for black victims. The inter-racial assault may have had a threatening effect on some subjects, and this threat may have been heightened in the case of the attractive black victim.

Further study is clearly warranted in this area first to confirm the replicability of the victim race effects on responsibility attributions and then to determine the most compelling explanation for why subjects are influenced by victim race. Preparations are in progress for a follow-up study in which new victim photographs and additional racial/ethnic categories will be used to study subjects’ perceptions of a victimization event.
Table 1. Manipulation Check On Victim Attractiveness

How would you rate the physical attractiveness of the victim?

<table>
<thead>
<tr>
<th>Victim Attractiveness</th>
<th>n</th>
<th>mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>122</td>
<td>7.80</td>
<td>2.00</td>
</tr>
<tr>
<td>Low</td>
<td>96</td>
<td>3.69</td>
<td>2.07</td>
</tr>
</tbody>
</table>

Table 2. Mean scores for attributions of responsibility to the defendant as a function of victim race and attractiveness

How much responsibility would you attribute to the defendant for the crime?

<table>
<thead>
<tr>
<th>Victim Race</th>
<th>Victim Attractiveness</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>x</td>
<td>10.12</td>
<td>9.85</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>1.98</td>
<td>1.65</td>
</tr>
<tr>
<td>Black</td>
<td>x</td>
<td>9.90</td>
<td>10.62</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>2.40</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Victims and offenders

Table 3. Mean sentencing recommendations for guilty defendant

Assuming that the defendant would be found guilty, what sentence would you recommend?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Victim Attractiveness**

<table>
<thead>
<tr>
<th>Victim Race</th>
<th>High</th>
<th>Low</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>x</td>
<td>8.04</td>
<td>6.56</td>
<td>7.73</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>2.57</td>
<td>2.50</td>
<td>3.03</td>
</tr>
<tr>
<td>Black</td>
<td>x</td>
<td>8.22</td>
<td>8.90</td>
<td>7.36</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>2.92</td>
<td>2.57</td>
<td>2.51</td>
</tr>
</tbody>
</table>

REFERENCES


Victims and offenders

INTRODUCTION

Lying is a fact of everyday life. Adults tell on average 1.5 lies a day (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996), and children are already lying by the age of four years (Lewis, Stanger, & Sullivan, 1989; Newton, Reddy, & Bull, 2000; Vrij, 2002a). People tell outright lies (lies where the information conveyed is completely different from what the deceiver believes to be the truth (“I was at home watching television” is an outright lie when the person in fact was burgling somebody’s house)); they exaggerate (embellishing remorse for committing a crime during a police interview or presenting themselves to be more diligent than in fact is the case during a job interview, etc.); or tell subtle lies. Subtle lying involves telling literal truths that are designed to mislead. The then President of the USA, Bill Clinton, was telling such a lie when he said to the American people that he “did not have sexual relations with that woman, Miss Lewinsky.” The lie was subtle, because the statement implied that nothing of a sexual nature had happened between the two of them, whereas he was relying on the narrower definition that they did not have actual sexual intercourse. Another method of subtle lying involves concealing information by evading the question or omitting relevant details. The passenger who tells a customs officer what he has in his luggage is concealing information when he deliberately fails to mention that he is also carrying illegal drugs.

[1] The studies, which are described in more detail (Tables 1, 3 and 4), were sponsored by the Economic and Social Research Council (grants R00429734727 and R000222820).
1.2 | REASONS TO LIE

People are motivated to lie for different reasons (DePaulo et al., 1996; Vrij, 2002b). They lie to obtain personal advantage (for example, not mentioning all the shortcomings when trying to sell their house); to avoid punishment (such as when guilty suspects deny involvement in the crime); to give a positive impression to others or to protect themselves from embarrassment or disapproval (for example, not admitting to being homosexual); to make others appear better, or for another person’s benefit (an innocent mother may tell the police that she has committed a crime in order to save her guilty son from a conviction); and for the sake of social relationships (so-called ‘social lies’, Vrij (2000a)). Goffman (1959) pointed out that life is like a theatre and that people often behave as actors and put on a show. Conversations could become awkward and unnecessarily rude without the social lubrication of deception, and social interactions can easily become disturbed when people tell each other the truth all the time (“I didn’t like the food you prepared”, “I don’t like the present you gave me”, and so on).

2 | INDIVIDUAL AND SITUATIONAL DIFFERENCES IN DECEIT

The frequency of lying depends on several factors, such as the personality of the liar, the situation in which the lie is told and to whom the lie is told. Regarding personality, extraverts lie more often than introverts (Kashy & DePaulo, 1996). However, instead of labelling introverts as ‘honest’, we often label them as ‘socially awkward’, perhaps because they are honest and tell fewer social lies. Although no gender differences have been found in the frequency of lying, men and women tell different types of lies, with women telling more social lies (DePaulo & Bell, 1996; DePaulo et al., 1996). This might be a reason why both men and women appreciate conversations with women more than conversations with men (DePaulo, Epstein, & Wyer, 1993; Reis, Senchak, & Solomon, 1985). Gender differences also emerged in the types of lies that men and women told during a date (Eyre, Read, & Milstein, 1997; Tooke & Camire, 1991). Women more frequently engaged in deceptive acts to improve their physical appearance (e.g. ‘sucking in’ their stomach when around members of the other sex), whereas men tended to feign their earning potential (e.g. misleading members of the opposite sex about their career success).

Regarding situation, Rowatt, Cunningham and Druen (1998) found that 90% of participants admitted to being willing to tell a lie to a prospective date. Also DePaulo and Kashy (1998) found that people lied relatively often to their romantic partners in the early stages of their relationship (once in every three interactions). One possible explanation is that people wondered whether their ‘true self’ was lovable enough to attract and keep these partners, and they therefore presented themselves as they wished they were instead of how they actually were (DePaulo & Kashy, 1998). Robinson, Shepherd and Heywood (1998) interviewed
undergraduate students, of whom 83% said they would lie in order to get a job. These students said that it was wrong to lie to best friends, but they saw nothing wrong in lying if this secured a job. They also thought that employers expected candidates to exaggerate qualities when applying.

Regarding to whom the lie is told, DePaulo and Kashy (1998) found that the lowest rate of lying occurred in conversations with spouses, while the highest rate occurred with strangers. However, the results clearly demonstrated that deception occurs in all types of close personal relationships. Although participants said they were predominantly honest in social interactions with their spouses, lies still occurred in nearly one out of every ten interactions they had with them. Many of those lies were minor, but interactions with spouses are also the domain of serious lies. When people were asked to describe the most serious lie they have ever told to someone else, they overwhelmingly reported that the target of these lies were romantic partners (Anderson, Ansfield, & DePaulo, 1999). These lies were often told to cover serious issues, such as infidelities, and were told to save the relationship. Sometimes spouses believe that the truth cannot be told without threatening the relationship. In such instances, they may decide that telling a lie is preferable. They perhaps do so reluctantly. They often feel uncomfortable while lying to their spouse (DePaulo & Kashy, 1998), but it is in their view the best option they have in those circumstances.

One reason why people lie less to their romantic partners (and also to friends) than to strangers is that they have a desire to be honest to people they feel close to, but there are also other reasons (Anderson, Ansfield et al., 1999). The fact that friends and partners know more about us limits the topics that are suitable or ‘safe’ to lie about. We can try to impress strangers at a cocktail party by exaggerating our culinary skills but this is ineffective with friends who have experienced our cooking. So, we might lie less to people we know well because we may think that we will not get away with deceit.

People tell different lies to their romantic partners than to people they know less well. Metts (1989) found that people are much less likely to tell outright lies to their romantic partners. They believe that the risks are too high and that the partner will eventually find out that they are lying. Moreover, they expect problems as soon as the outright lie is detected, because a liar may find it difficult to justify the fact that they lied to their partner, without appearing untrustworthy. Lies told to spouses are therefore usually subtle lies such as concealments. These are usually difficult to detect, because the liar does not reveal information that can be checked. The lie is also easier to justify when the truth emerges. The liar could just say that s/he simply forgot to mention a specific detail, or did not consider it important enough to bring up, and so on.

Although people tend to lie less to those with whom they feel close, there are exceptions. For example, a consistent finding is that college students lie frequently to their mothers (Backbier & Sieswerda, 1997; DePaulo & Kashy, 1998;
DePaulo and Kashy (1998) found that students lied in almost half of the conversations they had with their mothers. Perhaps they are still dependent on their mothers (for example, financially) and sometimes have to lie in order to secure monetary resources. Another explanation is that they do not want their mothers to worry or be disapproving of them. They therefore tell their mothers that they do not drink much beer, that they attend all lectures, that they study hard for their exams and that they regularly clean their room.

The fact that lying is a daily life event that commences at an early age, implies that the vast majority of people are experienced liars. Therefore, it is reasonable to suggest that people are skillful liars and that detecting deceit is difficult. Research, outlined below, has consistently supported this view.

3 | DECEPTIVE BEHAVIOUR

Two recent reviews, including more than a hundred studies examining the relationship between nonverbal behaviour and deception, have convincingly demonstrated that a typical nonverbal response to deception does not exist (DePaulo, Lindsay, Malone, Muhlenbruck, Charlton, & Cooper, in press; Vrij, 2000a). In other words, there is no giveaway clue like Pinocchio’s growing nose. However, some behaviours are more likely to occur during deception than others, depending on three processes that a liar may experience: emotion, content complexity and attempted behavioural control (DePaulo, Stone, & Lassiter, 1985a; Zuckerman, DePaulo, & Rosenthal, 1981; Vrij, 2000a). The mere fact that a person is lying will not result in any particular behaviour, but liars might be nervous (emotional), might have to think hard to come up with a plausible and convincing answer (content complexity) and might try to control their behaviour in order to give a credible impression (attempted behavioural control). Each of these three processes may elicit specific behavioural responses. The distinction between them is artificial. Lies may well feature all three aspects, and the three processes should not be considered as opposing camps.

3.1 | EMOTIONS

Telling a lie might evoke emotions. The three most common types of emotion associated with deceit are guilt, fear and duping delight (Ekman, 1992). A liar might feel guilty because (s)he is lying, might be afraid of getting caught, or might be excited about having the opportunity to fool someone. The strength of these emotions depends on the personality of the liar and on the circumstances.
under which the lie takes place (Ekman, 1992; Vrij, 2000a). Guilt, fear and excitement may influence a liar’s behaviour. Guilt might result in gaze aversion because the liar does not feel able to look the deceived person straight in the eye while telling a lie. Fear and excitement might result in signs of arousal, such as an increase in limb movements (movements of arms, hands, fingers, legs and feet), an increase in speech fillers (pauses in speech filled with ‘ah’, ‘um’, ‘er’ and so on), speech errors (words and/or sentence repetition, sentence change, sentence incompletions, slips of the tongue, and so on), facial emotional expressions (expressions of fear, anger, disgust etc.) or a higher pitched voice.

3.2 | CONTENT COMPLEXITY

Sometimes lying can be difficult, as liars have to think of plausible answers, avoid contradicting themselves, tell a lie that is consistent with everything that the observer knows or might find out, and avoid making slips of the tongue. Moreover, liars have to remember what they have said, so that they can be consistent when asked to repeat their story. People engaged in cognitively complex tasks make more speech fillers and speech errors, pause more, and wait longer before giving an answer (Goldman-Eisler, 1968). Cognitive complexity also leads to fewer limb movements and to more gaze aversion. The decrease in limb movements is due to the fact that a greater cognitive load results in a neglect of body language, reducing overall animation (Ekman & Friesen, 1972). Gaze aversion (usually to a motionless point) occurs because looking at the conversation partner distracts from thought. It is easy to examine the impact of content complexity on movements and gaze aversion. Ask people what they ate three days ago, and observe their behaviour while they try to remember. Most people will look away and will sit still while thinking about the answer.

3.3 | ATTEMPTED BEHAVIOURAL CONTROL

So far, the predictions of how liars behave have been straightforward. A liar may experience emotions and/or may find it difficult to lie, and this will result in behavioural signs of emotion and content complexity. However, the situation is more complicated than this. Liars may worry that several cues will give their lies away, and therefore will try to suppress such signs and might engage in impression management (Krauss, 1981) in order to avoid getting caught. That is, they may try to give a convincing impression. Most people lie less frequently than they tell the truth. This makes lying a special event that merits special attention. While lying, people may worry about the impression they make on others and may be keen on making an honest impression, perhaps even more so than when they are telling the truth. Someone who smuggles contraband is probably keener to make an honest impression on customs officers than someone who is not smuggling. It will not harm the non-smuggler when a customs officer asks her to open her suitcase. She might be annoyed about the time delay and inconvenience it cau-
Victims and offenders

Making an honest and convincing impression is not easy. It requires suppressing nerves effectively, masking evidence of heightened cognitive load, knowledge of how an honest person normally behaves and the ability to show the behaviour that is required. Hocking and Leathers (1980) argued that liars’ attempts to control their behaviour will focus on the cultural stereotype of deceptive behaviour. For example, if there is a widespread belief that liars look away, increase their movements and stutter, then liars will try to maintain eye contact, refrain from making too many movements and speak fluently. Eye contact should be easier to control than body movements and speech fillers and speech errors (Ekman & Friesen, 1974). The face is important in the exchange of information. For example, via facial expressions people can express whether they are interested in someone’s conversation, and whether they feel happy or sad (Ekman, 1992). The great communicative potential of the face means that people are well-practised at using and therefore controlling it. The body, on the other hand, is a channel that may not be salient in communication and is less often attended to and reacted to by others. We are therefore less practised in making movements and less good at controlling them. It may well be the case that, when controlling their behaviour, liars exhibit a pattern of behaviour that will appear planned, rehearsed, and lacking in spontaneity. For example, liars may believe that movements will give their lies away, and will therefore move very deliberately and tend to avoid any movements that are not strictly essential. This will result in an unusual degree of rigidity and inhibition, because people normally make movements that are not essential (DePaulo & Kirkendol, 1989).

Like most movements, speech fillers and speech errors are usually made unintentionally and are not important in the exchange of information. We therefore may assume that people do not often practise controlling these behaviours, and are not very good at controlling them. It is likely that liars will think that the use of speech fillers and speech errors sound suspect (Akehurst, Köhnken, Vrij, & Bull, 1996; Vrij & Semin, 1996). Therefore liars will try to avoid making such nonfluencies. This, however, may result in a speech pattern, which sounds unusually smooth, as it is normal for most people to make some errors in speech (Vrij & Heaven, 1999).

Another possible cue as a result of inadequate control of behaviour is that performances may look flat due to a lack of involvement (DePaulo et al., in press). An artist, who applies for a job as salesperson because he needs the money, may not look enthusiastic enough about the job opportunity during the selection interview. A mother who punishes her child for wrongdoing might not look sincere enough if she, in fact, was amused by the trick played on her.

All three processes may occur at the same time. That is, liars could be nervous, experiencing cognitive load, and trying to control themselves all at the same time.
Which of these processes is most prevalent depends on the type of lie. Liars will be more nervous when the stakes (negative consequences of getting caught and positive consequences of getting away with the lie) are high, hence, nervous behaviours are most likely to occur in high-stake lies. Liars have to think harder when the lie is complicated, therefore indicators of cognitive load are more likely to occur in complicated lies than in easy lies. Liars who are motivated to avoid getting caught may try harder to make an honest impression than those who are less motivated. Therefore, attempts to control behaviour may especially occur in motivated liars.

3.4 | POST HOC EXPLANATIONS

The three processes are hypothetical and are typically introduced post hoc to explain nonverbal differences between liars and truth tellers. Apart from this ‘three factor model’, other theoretical models for explaining nonverbal cues to deception are given as well in the deception literature (Buller & Burgoon, 1996; DePaulo et al., in press; Ekman, 1992; Ekman & Friesen, 1969). See DePaulo et al. (in press) for a description of each of these theoretical models. By discussing only the ‘three factor model’ we do not suggest that this theoretical model is superior to the other theoretical models. However, there is evidence that liars actually experience the three processes described in the three factor, whereas similar studies have not been carried out regarding the other theoretical models. In Vrij, Semin, and Bull’s (1996) experiment, participants were asked either to lie or to tell the truth. Afterwards they were asked to what extent they had experienced the three processes. Results showed that liars experienced all three processes significantly more than truth tellers. Vrij, Edward and Bull (2001b) found individual differences in experiencing these processes. For example, a negative correlation was found between being good at acting and having to think hard while lying. Although these studies were correlational studies, the relationship between the three processes and lying is more likely to be causal: They are the consequence of being engaged in lying.

3.5 | HOW DO LIARS BEHAVE?

The literature reviews (DePaulo et al., in press; Vrij, 2000a) have revealed that four behaviours in particular are more likely to occur during deception than while telling the truth: A higher pitched voice, an increase in speech errors (in particular an increase in word and phrase repetitions (DePaulo et al., in press)), a decrease in illustrators (hand and arm movements designed to modify and/or supplement what is being said verbally), and a decrease in hand/finger movements (movements of hands or fingers without moving the arms). These findings provide support for all three processes. The increase in pitch of voice might be the result of arousal experienced by liars (Ekman, Friesen, & Scherer, 1976). However, differences in pitch between liars and truth tellers are usually very small and therefore only detectable with sophisticated equip-
Victims and offenders

The increase in speech errors might be the result of liars having to think hard about their answer (there is some evidence that the increase in speech errors particularly occurs in complex lies (Vrij, 2000a)). Alternatively, the increase in speech errors might be caused by nervousness. The decrease in illustrators and hand/finger movements during deception might be the result of lie complexity: Perhaps liars experience cognitive load, resulting in a neglect of body language. Another explanation is that liars, in an effort to make an honest impression, move very deliberately and try to avoid those movements which are not strictly essential, resulting in an unusual degree of rigidity and inhibition.

Perhaps a surprising finding is that liars do not seem to show clear signs of nervousness, such as gaze aversion and fidgeting. At least in white Western cultures, there is a strong stereotypical belief amongst observers, including professional lie catchers, that liars look away and make grooming gestures (Akehurst et al., 1996; Vrij & Semin, 1996).

4 | REASONS FOR THE ABSENCE OF BEHAVIOURAL CUES TO DECEIT

Scholars have provided several explanations why nonverbal indicators of deceit are often not found in deception research (Vrij, 2000a), including the following three important reasons.

4.1 | LACK OF DETAIL IN SCORING SYSTEMS

First, perhaps some indicators are overlooked by researchers, because the scoring systems they use to measure the occurrence of behaviours are not appropriate to reveal indicators of deceit. This is probably true regarding smiles. Literature reviews do not show a relationship between deception and smiling (DePaulo et al., in press; Vrij, 2000a). However, Ekman and colleagues discovered that smiles are related to deception when a distinction is made between felt and false smiles (Ekman, Friesen, & O’Sullivan, 1988). They found that truth tellers showed more felt smiles and liars more false smiles. Felt smiles include all smiles in which the person actually experiences a positive emotion and presumably would report that positive emotion. False smiles are deliberately made to convince another person that a positive emotion is felt whereas, in fact, it isn’t. Felt and false smiles involve the use of slightly different facial muscle actions and the skilled observer is able to spot these differences (Ekman, 1992).

Commercial companies have exploited this idea and brought several voice analysers on the market, which, they say, can be used to detect deceit. However, these analysers are not as accurate as many companies claim them to be. See Vrij (2000a) for problems lie detectors face when using equipment that measures physiological responses.
4.2 | THREE PROCESSES

Second, truth tellers may show similar behaviour to liars, because they too may experience emotions, may have to think hard or may try to control themselves. For example, innocent (truthful) suspects might be anxious during police interviews because they might be worried that they will not be believed by a police detective (Ofshe & Leo, 1997). They then may show the same nervous behaviours as guilty suspects who are afraid of being caught, because the ‘fear of not being believed when truth telling’ and ‘fear of getting caught in a lie’ will produce the same behaviours (Bond & Fahey, 1987).

4.3 | RAISING THE STAKES

Third, the absence of significant findings might be the result of an artefact. Deception research has almost exclusively been conducted in university laboratories where participants (mostly college students) tell the truth or lie for the sake of the experiment. Perhaps in these laboratory studies the stakes are not high enough for the liar to elicit clear cues to deception (Miller & Stiff, 1993).

In order to raise the stakes in laboratory experiments, participants have been offered money if they successfully get away with their lies (Vrij, 1995). In other studies, participants are told that they will be observed by a peer who will judge their sincerity (DePaulo, Stone, & Lassiter, 1985b). In an attempt to raise the stakes even further, participants in Frank and Ekman’s (1997) study were given the opportunity to ‘steal’ US$50. If they could convince the interviewer that they had not taken the money, they could keep all of it. If they took the money and the interviewer judged them as lying, they had to give the US$50 back and also lost their US$10 per hour participation fee. Moreover, some participants faced an additional punishment if they were found to be lying. They were told that they would have to sit on a cold, metal chair inside a cramped, darkened room labelled ominously XXX, where they would have to endure anything from 10 to 40 randomly sequenced 110-decibel starting blasts of white noise over the course of one hour.

A study like this should raise ethical concerns. Also, one might argue that the stakes in such a study are still not comparable with the stakes in some real life situations, such as during police interviews. Laboratory studies are not suitable for examining the responses in high-stake situations as raising the stakes to a comparable extent is not usually possible for ethical reasons. Therefore, the only way to investigate how liars behave in high-stake real life situations is to examine their behavioural responses in such situations. This has proven to be difficult and, as a result, behavioural examinations of real life high-stake situations are virtually non-existent. Researchers face three problems in particular.
5 | DIFFICULTIES IN EXAMINING REAL LIFE HIGH-STAKE SITUATIONS

First, researchers have difficulty in obtaining appropriate video footage. For example, it is only relatively recently that interviews with suspects have been videotaped in a few constabularies in England and Wales and even when they are videotaped, researchers are rarely given permission to analyse those videotaped interviews.

Second, it is often difficult in real life cases to establish the so-called ground truth, that is, to obtain conclusive evidence that the person is lying or telling the truth. For example, when former US President Bill Clinton testified before the grand jury in 1998 about his alleged sexual affair with Monica Lewinsky, he showed distinctive behaviour during parts of the interview, in particular when asked about Betty Currie (Vrij, 1998, 2002b). Betty Currie (who was Clinton’s personal secretary) went to Monica Lewinsky’s home to collect the presents she had received from Clinton. The question was whether or not Clinton instructed her to do this. The question was asked several times throughout the interview, and each time Clinton denied that he had given Betty Currie any instructions. During these denials Clinton sat very still and looked straight into the camera. However, what is the ground truth in this case: Did Clinton ask Betty Currie to take away the presents or not? We simply don’t know.

Iraq’s President Saddam Hussein was interviewed by the journalist Peter Arnett from Cable News Network (CNN) during the Gulf War (1991). The interview was broadcast on CNN. In order to avoid destruction by Allied forces, Iraqi planes were sent to neighbouring Iran. Arnett discussed the Iraqi planes landing in Iran and whether they would return to Iraq for use. Hussein told Arnett, amongst others, that: “There isn’t one single Islamic country that is not on our side in this battle.” Davis and Hadiks (1995), who observed and scored Hussein’s behaviour during the CNN interview, noticed that throughout this part of the interview, Hussein’s behaviour was restricted and controlled. He slowly sat up, and markedly restricted and even stopped his gesticulations. Davis and Hadiks believe that at that moment Hussein was fabricating an answer. However, there is no conclusive evidence for this.

Third, to examine cues to deception, a useful method might be to compare the response under investigation with a response the lie detector knows to be truthful (so-called baseline method). However, the two situations (situation under investigation and responses while telling the truth) should be comparable. For example, Clinton showed specific behaviour when asked about Betty Currie. Why did he do that? Because he was lying? Or did this particular question trigger a specific behavioural response? This latter explanation cannot be ruled out. Davis and Hadiks’ (1995) observations revealed that Hussein used a variety of hand and arm movements, and that he made specific illustrators when discussing specific issues. In summary, specific situations sometimes result in specific behaviours.
Lie detectors should be aware of this and should compare situations that are similar to avoid comparing apples with oranges. Unfortunately, during police interviews apple-orange comparisons are sometimes made (Moston & Engelberg, 1993). In those interviews, suspects’ behaviour during small talk conversations at the beginning of the interview is compared with their behaviour during the actual interrogation. Although police officers are advised to establish comparable truths in this way (Inbau, Reid, & Buckley, 1986), this is an inaccurate comparison. Small talk and the actual investigation are different situations. Not surprisingly, both guilty and innocent people tend to show different behaviours during small talk and the actual interview (Vrij, 1995).

### 6 | HOW DO SUSPECTS BEHAVE DURING THEIR POLICE INTERVIEWS?

In the most comprehensive study of real life high-stake lies to date, Mann, Vrij, and Bull (2002b) examined the behaviour displayed by 16 suspects during their police interviews. The suspects were all being interviewed in connection with serious crimes such as murder, rape and arson. Regarding the ground truth, clips of video footage were selected where other sources (reliable witness statements and forensic evidence) provided evidence that the suspect lied or told the truth. In addition, for each suspect, truths and lies were chosen which were as comparable as possible in nature (for example, a suspect who gave a detailed description about how he had assisted in killing a person (truth), later denied any involvement in the crime (lie). Forensic evidence indisputably supported his original version). Table 1 shows the results of the study for the total sample (N = 16) and for male suspects only (N = 13).

Table 1. Differences between truthful and deceptive responses (Mann, Vrij, & Bull, 2002b)

<table>
<thead>
<tr>
<th></th>
<th>total sample (N - 16)</th>
<th>males only (N - 13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze aversion</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eye blinks</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>Head movements</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hand/arm movements</td>
<td>-</td>
<td>&lt;</td>
</tr>
<tr>
<td>Pauses</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>Speech disturbances</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

< decrease during deception
> increase during deception
- no difference
- speech fillers and speech errors combined
As can be seen in Table 1, results revealed that the suspects in these high-stake situations did not show clear stereotypical nervous behaviours such as gaze aversion, increased speech disturbances, or increased movements. In fact, they exhibited an increase in pauses and (male suspects) a decrease in hand and arm movements. This is more in agreement with the content complexity and attempted control approaches than with the emotional approach. The strongest evidence that content complexity affected suspects’ behaviour more than nervousness was the finding regarding eye blinks. Suspects made fewer eye blinks when they lied. Research has shown that nervousness results in an increase in eye blinking (Harrigan & O’Connell, 1996; Tecce, 1992), whereas increased cognitive load results in a decrease in eye blinking (Wallbott & Scherer, 1991).

The apparent predominance of cognitive load processes compared to emotional processes in suspect interviews is perhaps not surprising. Many suspects have had regular contact with the police. Therefore, they are probably familiar with police interviews, which might decrease their nervousness during those interviews. However, suspects in police interviews are often less intelligent than the average person (Gudjonsson, 1992). There is evidence that less intelligent people will have particular difficulty in inventing plausible and convincing stories (Ekman & Frank, 1993).

7 | DETECTING LIES

7.1 | LAY PERSONS’ ABILITY TO DETECT DECEIT

In scientific studies concerning the detection of deception, observers are typically given videotapes or audiotapes and asked to judge whether each of a number of people is lying or telling the truth. Statements from liars and truth tellers are usually taken from laboratory studies and rarely from real life situations. Vrij (2000a) examined the percentages of correct lie detection (the ‘accuracy rate’) across 37 studies. Included were studies in which judges were university students who tried to detect lies and truths told by people they did not know. The total accuracy rate was 56.6%, when 50% accuracy is expected by chance alone. (Simply guessing whether someone is lying or not allows for a 50% chance of being correct).

When accuracy at detecting lies was computed separately from accuracy at detecting truth, the results showed that observers were reasonably good at detecting truths (correctly judging that someone was telling the truth: 67% accuracy rate) but particularly poor at detecting lies (correctly judging that someone was lying: 44% accuracy rate). In fact, 44% is below the level of chance. In other words, people would be more accurate at detecting lies if they simply guessed!

The superior accuracy rate for truthful messages is caused by the truth bias: judges are more likely to consider that messages are truthful than that they are deceptive and, as a result, truthful messages are identified with more accuracy than are deceptive ones. There are four possible explanations for the truth bias.
First, in daily life people are more often confronted with truthful than with deceptive statements, so they are therefore more inclined to assume that the behaviour they observe is honest (the so-called availability heuristic, O’Sullivan, Ekman, & Friesen, 1988). Second, social conversation rules prevent people from displaying suspicion. A person will very quickly become irritated if their conversation partner questions everything that is being said. Unfortunately, it is often necessary to challenge what the person is saying and ask for more information in order to detect deceit (Vrij, 2000a). Third, the truth bias is the result of the incorrect stereotypical views people have about how liars behave. For example, most people expect liars to behave nervously (Akehurst et al., 1996; Vrij & Semin, 1996). Since many liars do not show nervous behaviours, observers who look for cues of nervousness to detect deceit are inclined to judge many messages as truthful. Fourth, people may be unsure as to whether deception in fact is occurring. Given this uncertainty, the safest and most polite strategy may be to believe what is overtly expressed (DePaulo, Jordan, Irvine, & Laser, 1982).

7.2 | PROFESSIONAL LIE CATCHERS’ ABILITY TO DETECT DECEIT

It could be argued that university students do not habitually detect deception. Perhaps professional lie catchers, such as police officers or customs officers, would obtain higher accuracy rates than laypersons. It might be that their experiences at interviewing people and catching liars has a positive influence on their skills to detect deceit. In some studies professional lie catchers participated as judges. Table 2 provides an overview of these studies.

Table 2. Accuracy scores of professional lie catchers

<table>
<thead>
<tr>
<th>Source</th>
<th>Truth</th>
<th>Lie</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>DePaulo and Pfeifer (1986) (federal law enforcement)</td>
<td>64%a</td>
<td>42%a</td>
<td>53%</td>
</tr>
<tr>
<td>Ekman &amp; O’Sullivan (1991) (Secret Service)</td>
<td></td>
<td></td>
<td>56%</td>
</tr>
<tr>
<td>Ekman &amp; O’Sullivan (1991) (federal polygraphers)</td>
<td></td>
<td></td>
<td>56%</td>
</tr>
<tr>
<td>Ekman &amp; O’Sullivan (1991) (police officers)</td>
<td></td>
<td></td>
<td>56%</td>
</tr>
<tr>
<td>Ekman, O’Sullivan, &amp; Frank (1999) (CIA)</td>
<td>66%</td>
<td>80%</td>
<td>73%</td>
</tr>
<tr>
<td>Ekman, O’Sullivan, &amp; Frank (1999) (sheriffs)</td>
<td>56%</td>
<td>78%</td>
<td>67%</td>
</tr>
<tr>
<td>Ekman, O’Sullivan, Frank (1999) (law enforcement)</td>
<td>54%</td>
<td>48%</td>
<td>51%</td>
</tr>
<tr>
<td>Köhnken (1987) (police officers)</td>
<td>58%</td>
<td>31%</td>
<td>45%</td>
</tr>
<tr>
<td>Porter, Woodworth, &amp; Birt (2000) (parole officers)</td>
<td>20%</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>Vrij (1993) (police detectives)</td>
<td>51%</td>
<td>46%</td>
<td>49%</td>
</tr>
<tr>
<td>Vrij &amp; Graham (1997) (police officers)</td>
<td></td>
<td></td>
<td>54%</td>
</tr>
<tr>
<td>Vrij &amp; Mann (2001a) (police officers)</td>
<td>70%</td>
<td>57%</td>
<td>64%</td>
</tr>
<tr>
<td>Vrij &amp; Mann (2001b) (police officers)</td>
<td></td>
<td></td>
<td>51%</td>
</tr>
</tbody>
</table>

a Experienced and inexperienced officers together
Four findings emerged from these studies. First, not many studies have been conducted with professional lie catchers, probably due to the fact that professional lie catchers are sometimes reluctant to participate in lie detection studies or do not wish the outcomes to be published. Second, most accuracy rates were similar to the accuracy rates found in studies with college students as observers, suggesting that professional lie catchers are no better than laypersons in detecting deceit. DePaulo and Pfeifer (1986), Ekman and O’Sullivan (1991) and Vrij and Graham (1997) directly tested this idea by including both laypersons and professional lie catchers as observers in their experiments. DePaulo and Pfeifer (1986) and Vrij and Graham (1997) found that police officers were as (un)succesful as university students in detecting deception. Ekman and O’Sullivan (1991) found that police officers and polygraph examiners obtained similar accuracy rates to university students, whereas members of the Secret Service were better at detecting lies than university students. Third, some groups seem to be better than others. Ekman’s research has shown that members of the Secret Service, CIA and sheriffs were better lie detectors than other divisions of police officers. Fourth, the truth-bias, consistently found in studies with students as observers, is much less profound or perhaps even lacking in studies with professional lie catchers. It might be that their jobs make professional lie catchers more suspicious.

Moreover, DePaulo and Pfeifer (1986) investigated how confident observers were in the decisions they made. They found that police officers were more confident than university students, which suggests that being a professional lie catcher may increase self-confidence in ability to detect deceit, but does not increase accuracy. The tendency to be overconfident is not restricted to police officers. It is common amongst many different groups of professionals (Allwood & Granhag, 1997).

8 | REASONS WHY PEOPLE ARE POOR LIE DETECTORS

There are numerous reasons why people are generally poor at detecting deceit (Vrij, 2000a), some of the main reasons, ‘Ostrich-approach’, Othello-error’, ‘incorrect beliefs’ and ‘low-stake lies’ are mentioned here.

8.1 | OSTRICH-APPROACH

Some lies remain undetected because observers do not want to detect a lie, because it is not in their best interests to learn the truth. People generally appreciate the compliments made by others about their body shape, their hairstyle, the way they are dressed, their achievements, and so on. So why bother with trying to discover whether those compliments are spoken in truth? After the scandal with Lewinsky broke, President Clinton told his aides in the White
House that he did not have a sexual relationship with Monica Lewinsky. Erskine Bowles, the then White House Chief of Staff, was more than willing to believe him. This is how he described that moment to the grand jury: “All I can tell you is: This guy who I’ve worked for looked me in the eye and said he did not have sexual relationships with her. And if I didn’t believe him, I couldn’t stay. So I believe him.”

8.2 | OTHELLO-ERROR

The fact that truth tellers and liars might have the same experiences and therefore behave similarly hampers lie detection. As mentioned before, both guilty and innocent suspects might be afraid during police interviews: guilty suspects because they are afraid of getting caught, and innocent suspects because they are afraid that they won’t be believed. Therefore, because of that fear, guilty and innocent suspects may show the same nervous behaviours. This puts lie detectors in a difficult position. Should the signs of fear be interpreted as a sign of guilt or as a sign of innocence? The behaviour doesn’t provide the answer. Unfortunately, lie detectors seem not to be fully aware of this and are often inclined to interpret signs of nervousness as deceptive even when they are displayed by truth tellers. Ekman (1992) labelled this phenomenon the *Othello error*, after Shakespeare’s play. Othello falsely accuses Desdemona (his wife) of infidelity. He tells her to confess since he is going to kill her for her treachery. Desdemona asks Cassio (her alleged lover) to be called so that he can testify her innocence. Othello tells her that he has already murdered Cassio. Realising that she cannot prove her innocence, Desdemona reacts with an emotional outburst. Othello misinterprets this outburst as a sign of her infidelity.

8.3 | INCORRECT BELIEFS

Observers seem to have incorrect beliefs about how liars behave. Vrij (2000a) reviewed studies examining people’s beliefs about deceptive behaviour. These studies were carried out in various countries (although mainly in the Western world with Western white participants), including the United States, the United Kingdom, Germany, and the Netherlands, and with a variety of observers, including laypersons, police officers and customs officers. Despite the variety of locations and observers, the findings were highly similar. It appears that there is common belief, at least amongst Western white people, about how liars behave. Results showed that observers associate deception with a high-pitched voice, many speech fillers and speech errors, a slow speech rate, a long latency period (period of silence between question and answer), many pauses, gaze aversion, a lot of smiling, and an increase in movements. Many of these behaviours are indicators of nervousness. Apparently, the stereotypical belief is that liars are nervous and will behave accordingly. As we saw earlier, most of these behaviours are not related to deception (such as gaze aversion) or are related to deception in a diffe-
rent way (for example, illustrators and hand/finger movements tend to decrease during deception rather than increase).

People look at the wrong cues for various reasons. Vrij, Edward, and Bull (2001a) and Vrij et al. (1996) investigated participants’ behaviour while lying and truth telling. They also asked the participants afterwards to indicate how they thought they behaved when they lied and when they were telling the truth. Results showed that participants had poor insight into their own behaviour and thought that they responded more stereotypically while lying (showing gaze aversion, an increase in movements, and so on) than they in fact did. In other words, it seems that during lie detection observers look for cues they mistakenly believe they themselves show while lying.

Another reason is that people, including police officers, are taught the wrong cues (Gordon & Fleisher, 2002; Hess, 1997; Inbau et al., 1986). In their influential manual Criminal interrogation and confessions, Inbau et al. (1986) describe in detail how, in their view, liars behave. This includes behaviours such as showing gaze aversion, displaying unnatural posture changes, fidgeting and placing their hand over their mouth or eyes when speaking. They based their view on their extensive experience with interviewing suspects. However, none of these behaviours are found to be reliably related to deception in deception research. Neither do Inbau and his colleagues provide any empirical evidence for their claims.

Finally, it might be that children show those stereotypical behaviours when they lie. Perhaps, adults form their views about how liars behave by watching (their own) children’s responses, and assume that adults will show similar behaviours during deception. The core question then is: Do children look away and fidget when they lie? The studies, which have been published to date, do not suggest that they do (Vrij, 2002a). Vrij, Akehurst, Soukara and Bull (in press) were the first researchers to directly compare adults and children’s nonverbal responses during deceit. Although they found large differences in behaviour, for example, children (5-6 year-olds) made almost twice as many movements as adults (college students), deceptive cues were remarkably similar as can be seen in Table 3.

In both college students and 5-6 year-olds, deception was associated with a decrease in hand and finger movements. Stereotypical cues, such as gaze aversion and fidgeting, did not reveal deception in either group of participants.
Table 3. Differences between truthful and deceptive responses for college students and 5-6-year olds (Vrij, Akehurst, Soukara, & Bull, in press)

<table>
<thead>
<tr>
<th></th>
<th>5-6-year-olds</th>
<th>college students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze aversion</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>fidgeting</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>illustrators</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>hand/finger movements</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>Pauses</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Speech fillers</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>speech errors</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

< decrease during deception
- no difference

8.4 | LOW-STAKE LIES

Similar to studies examining deceptive behaviour, the findings of studies examining people’s ability to detect deceit might be the result of an artefact. There are clear differences between lie detection in scientific deception studies and in some real life situations. In detection of deception studies, observers are exposed to communicators who tell the truth or lie for the sake of the experiment. For those communicators, the stakes are probably much lower than, for example, for suspects in police interviews. Perhaps lie detection is much more difficult in low-stake situations as the three processes which might elicit behavioural responses to deceit (emotions, cognitive complexity, and attempted behavioural control) will be less profound in low-stake situations. Indeed, in a series of experiments in which the stakes were manipulated (although the stakes were never really high), it was found that high-stake lies were easier to detect than low-stake lies (DePaulo, Kirkendol, Tang, & O’Brien, 1988; DePaulo, Lanier, & Davis, 1983; DePaulo, LeMay, & Epstein, 1991; DePaulo et al., 1989b; Lane & DePaulo, 1999; Vrij, 2000b; Vrij, Harden, Terry, Edward, & Bull, 2001). In other words, detecting lies in low-stake situations is different from detecting lies in high-stake situations and therefore police officers’ true ability to detect deceit can probably only be examined when they are exposed to truths and lies told in real police interviews.

9 | DETECTING LIES IN A REAL LIFE HIGH-STAKE SITUATION

We were the first researchers to have conducted a lie detection experiment using real life high-stake materials (Mann, Vrij, & Bull, 2002a). We showed 99 police officers a total of 54 video clips in which suspects told truths and lies during their police interviews. After each clip the police officers were requested to indicate
Victims and offenders

whether the suspect was truthful or not. The truthful and deceptive clips were derived from Mann et al.'s (2002b) study, from which the findings have already been discussed in Table 1. None of the sample of police officers belonged to one of the specific groups, which have been identified by Ekman and his colleagues as being superior lie detectors (see Table 2).

The study revealed accuracy rates that were higher than generally found in previous studies. The total accuracy rate was 64.85%, with a 63.61% truth accuracy rate and a 66.16% lie accuracy rate. All these accuracy rates were significantly higher than the 50% level of chance. Moreover, both the total accuracy rate and the lie accuracy rates were higher than the accuracy rates that were found in most previous studies (see Vrij (2000a) for a review). In other words, ordinary police officers seem to be better at detecting truths and lies than was previously suggested. Although the total and lie accuracy rates were significantly higher than the total and lie accuracy scores obtained by laypersons (mostly college students) in previous research, it cannot be concluded that police officers are actually better lie detectors than laypersons. Laypersons were not included in this study, and perhaps they would have achieved similar accuracy rates as police officers if they had participated.

We also found individual differences, with some police officers performing better at the lie detecting task than others. It is relevant to discover who the good lie detectors are. Before and after the lie detection task, participants were asked which cues they believed indicated deceit, and possible relationships between those answers and their performance on the lie detection task were explored. Relationships between the cues people believe indicate deception and their performance on lie detection tasks have previously been explored by other researchers. We sought to replicate these previous findings by re-running their analyses on our data. The results are shown in Table 4.

Table 4. Relationships between cues people believe indicate deception and their performance on lie detection tasks (Mann, Vrij, & Bull, 2002a)

<table>
<thead>
<tr>
<th>Cue categories</th>
<th>they found</th>
<th>we found</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>truth</td>
<td>lie</td>
</tr>
<tr>
<td>Inbau, Reid, &amp; Buckley (1986)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nonverbal</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Ekman &amp; O’Sullivan (1991)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>speech</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>nonverbal behaviour</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>both</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Feeley &amp; Young (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>verbal</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>demeanour</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
First, we tested the usefulness of Inbau et al.’s (1986) nonverbal cues to detect deceit. Although they did not report data themselves, they have indicated that a liar’s posture, putting the hand over the mouth, fidgeting and gaze behaviour indicate deceit. We found that the more participants said they relied on these cues, the lower their truth accuracy and total accuracy were. In other words, looking at Inbau’s cues is counterproductive. Kassin and Fong (1999) obtained a similar finding. In their experiment, they taught half of their participants to look at the nonverbal cues mentioned by Inbau and colleagues as indicators of deceit. This group of participants actually performed worse on a subsequent lie detection task than a group of participants who were not informed about these nonverbal cues.

Ekman and O’Sullivan (1991) divided their participants into three groups. Those who mentioned only nonverbal cues, those who mentioned only verbal cues and those who mentioned both verbal and nonverbal cues. They found that those participants who mentioned both speech cues and nonverbal cues obtained the highest total accuracy. We could not replicate their finding on our data. We found that those who only mentioned speech cues obtained the highest accuracy score.

Feeley and Young (2000) clustered the cues mentioned by participants into four distinctive categories: verbal cues, demeanour cues, visual cues and vocal cues. They found that the more vocal cues (speech errors, speech fillers, pauses, voice) participants mentioned, the higher total accuracy they obtained. We could not replicate this finding with our data. We found that the more visual cues (gaze, smiles, movements, posture, etc.) the participants mentioned, the lower lie and total accuracy rates the participants obtained.
Anderson, DePaulo, Ansfield, Tickle, and Green’s (1999) findings were similar to the findings of Feeley and Young (2000). The more paralinguistic cues (identical to Feeley and Young’s (2000) ‘vocal’ category) the participants mentioned, the higher total accuracy score they obtained. As already mentioned, we could not replicate this finding, but we found that the more visual cues (Feeley and Young’s (2000) visual category and demeanour cues (defensive, confident, nervous) added) the participants mentioned, the lower their lie and total accuracy scores became.

Porter, Woodworth and Birt (2000) found that the more visual cues (identical to Feeley and Young’s (2000) category) the participants reported, the higher their truth and lie accuracy scores became. As already mentioned above, re-running that analysis on our data showed the opposite effect: The more visual cues the participants listed, the lower lie and total accuracy scores they received.

Finally, Vrij and Mann (2001a) found that those participants who mentioned gaze aversion and fidgeting as cues to deceit obtained the lowest total accuracy scores. This finding was replicated with our own data.

In summary, our data suggest that those participants who reported that visual cues, particularly gaze aversion and fidgeting, indicate deception were the worst lie detectors. We also found some evidence that those who reported that they look at speech cues to detect lies were the best lie detectors. However, when we compared these findings with the findings of previous studies, we could only replicate our own previous findings (Vrij & Mann, 2001a). How can we explain this lack of replication? One explanation is that the relationships between cues mentioned and accuracy are generally weak. Indeed, the significant correlations which are typically reported are low, they usually fall into the r = .20 to r = .30 range. Another explanation is that in different studies participants face completely different lie detection situations, and that therefore comparisons are difficult to make. For example, in most previous studies participants were requested to detect truths and lies in low-stake situations, whereas we exposed our participants to high-stake situations. The only previous study, which involved a high-stake situation, was Vrij and Mann’s (2001a) experiment in which participants were requested to detect the truths and lies told during a police interview by a man who was suspected of murder. We were able to replicate the findings of that study. Thirdly, perhaps lie detectors simply do not know where they look and so lie detection could just be a skill based on intuition. We obtained some support for this assumption, as good lie detectors reported relying significantly more often to on ‘gut feeling’ than poor lie detectors. Finally, it might be that good or bad performance on our lie detection task and on the lie detection tasks in previous studies was partly a matter of luck. In order to investigate this, several lie detection tasks need to be carried out with the same participants and consistency between the performances needs to be examined (a good lie detector can be defined as someone who has consistently high accuracy rates on different lie detection tasks). Do observers who consistently perform well look at different cues to the other lie detectors? Our data do not provide
TELLING AND DETECTING TRUE LIES

the opportunity to answer this question as only one lie detection task was conducted. To our knowledge, only Frank and Ekman (1997) exposed lie detectors to multiple lie detection tasks and they reported some kind of consistency between the lie detectors’ performances on the different tasks. They also reported that good lie detectors were better at spotting brief facial expressions of emotions than poor lie detectors. Such micro expressions have not been investigated in any of the other studies discussed here.

10 SUMMARY OF TELLING AND DETECTING REAL LIFE HIGH-STAKE LIES

Research examining (i) nonverbal responses while lying and (ii) people’s ability to detect deceit have limitations which make it difficult to generalise the findings to real life high-stake situations, such as police interviews. Our studies (Mann et al., 2002a, b) which examined the nonverbal responses of truth tellers and liars during police interviews and police officers’ ability to detect such truths and lies, revealed that (i) cues indicating that suspects experience cognitive load are the most likely cues to occur during deception in police interviews, and (ii) police officers are to some extent able to detect deceit during police interviews. However, erroneous decisions about the veracity of suspects did frequently occur, which is perhaps not surprising given the fact that suspects, and people in general, lie frequently and are therefore experienced liars.

REFERENCES


Victims and offenders


TELLING AND DETECTING TRUE LIES


Victims and offenders


Violence Risk Assessment

AN ANCHORED NARRATIVE APPROACH

Aldert VRIJ1, Samantha MANN

I | INTRODUCTION

Those who work in the field of violence risk assessment can be justifiably proud of their advances over the past two decades. The state of the discipline has progressed from a general attitude of nihilism, or “nothing works,” to one of optimism – not cautious or guarded optimism, but smile-on-the-face optimism. One major area of success is the identification of individual difference variables that discriminate repetitively violent people from those whose violence is rare, occasional, or episodic. Another is the demonstration that, using a variety of methods and procedures, human judges can make predictions of violence with reasonable accuracy – at least, reasonable at the group or aggregate level.

In this paper, however, I won’t dwell on successes. Instead, I will discuss some recent thoughts – or, more accurately, concerns – about the manner in which we are trying to understand, predict, and prevent violence. My general thesis is that we have progressed about as far as we can, relying solely on our current methodological tools: We are at, or at least we are nearing, the point of diminishing returns; we are straying dangerously close to a conceptual rut, and it’s time to try...
something new. Gerd Gigerenzer has pointed out that many advances in psychological theory were spurred by the development of new methodologies, and in particular statistical methods. He calls this the “tools-to-theories” heuristic (Gigerenzer, 2000). This is a twist on the old axiom, “When all you have is a hammer, everything looks like a nail.” In Gigerenzer’s version, not only does everything look like a nail to you, but also you start to think the universe is, in fact, made of nails. Having been persuaded by Gigerenzer on this point, I am of the opinion that further theoretical progress in the field of violence risk assessment awaits the development of new methodological tools. At the 2002 annual meeting of the European Association of Psychology and Law in Lisboa, Portugal, I discussed in some detail the limitations of actuarial risk assessment and then provided a summary of a possible alternative, which I referred to as the “anchored narrative” approach (Hart, in press). This year, I will do the opposite: Summarize the limitations of actuarial risk assessment, and then discuss in some detail the anchored narrative approach.

The thoughts I will share with you reflect collaborations and discussion with some cherished friends, including colleagues from Simon Fraser University (people such as Chris Webster, Randy Kropp, Ron Roesch, Ray Corrado, Derek Eaves, and Kelly Watt), the United States (John Monahan, Kirk Heilbrun, Joel Dvoskin, Dale McNeil, and Tom Litwack), and Europe (David Cooke, Henrik Belfrage, Martin Grann, and Caroline Logan). Of course, lest I be accused of minimization or denial, let me make it clear that I alone am responsible for any comment that you find trite, fallacious, or otherwise objectionable.

[2] The “conceptual rut” to which I refer is the period of that often follows scientific advances in which people are enamored with some new technique or procedure and spend their time testing it out rather than innovating. For example, with respect to violence risk, one could view the development of actuarial and structured professional judgement procedures in the 1990s as “advances.” Subsequent to the advent of instruments such as the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 1998) or the HCR-20 (Webster, Douglas, Eaves, & Hart, 1997), considerable time and effort has been devoted to evaluating their generalizability using methods either the same as or similar to those used in the original validation research: Do they work as well in Sweden or the Netherlands or Germany or the United States as they do in Canada (e.g., Grann, Belfrage, & Tengström, 2000)? As well with women as with men (e.g., Strand & Belfrage, 2001)? As well with adolescents as with adults (e.g., MacEachern, 2001)? As well with correctional offenders as with civil and forensic psychiatric patients (e.g., Belfrage, Fransson, & Strand, 2000)? Research of this sort yields critical information relevant to practice, but is intended to clarify the limitations of previous research findings rather than to generate new knowledge or theory.

[3] I located this axiom in numerous compendia, in which it was sometimes referred to as Baruch’s Law and phrased, “To the man with a hammer, every problem looks like a nail.” I was unable to determine the original source.
I have come to believe that the obstacles we face in understanding, assessing, and managing violence risk stem from the influence of predictionism, the belief that the universe and everything in it, including human behavior, can be described tolerably well in terms of stochastic models, axiom systems that include random variables (Bernstein, 1996; Gigerenzer et al., 1989; Jamieson, 2000; Oreskes, 2000). This has led to the rise of the "prediction enterprise" in modern society, a "dynamic social and political milieu" involving policy makers, institutions, and the public whose goal is to generate information that can guide social action (Pielke, Sarewitz, & Byerly, 2000, p. 362). In the field of violence risk, the predictionist view is that a few key variables – properly assessed, weighted, and combined – are sufficient to explain the occurrence of violence with satisfactory precision. For whatever reason, predictionism seems to be especially influential in North America.

Regardless of its origins and the reasons for its ascendance, predictionism now is with us. I will now summarize criticisms of predictionism I have discussed in various places over the past decade (Hart, 1996, 1998, 2002, in press). I hope to convince you in these next few minutes that predictive accuracy is grossly overrated as a criterion for evaluating our understanding of violence risk: Prediction is not central to policy-making; prediction is not central to the validation of scientific theory; and prediction pretends a degree of certainty that can never be attained in our field. All the criticisms come from within "hard sciences" like physics and mathematics, or from applied sciences like climatology and environmental toxicology; I promise not to utter the word "post-modern.")

[4] As I have noted previously, the leading figures associated with predictionism in the field of violence risk assessment are North American, people such as Paul Meehl, John Monahan, Hank Steadman, Vern Quinsey, Marnie Rice, and Grant Harris. This may have something to do with the historical insularity of North America – that is, insulation from its philosophical and moral ancestry. Although the dominant North American culture was established and subsequently reinvigorated on a near-continuous basis by Europeans, it has always been forward-looking. As Hegel noted, European intellectual tradition is deeply rooted in the past, aware that progress (social, scientific, or political) is a slow, painful, cumulative, and sometimes messy process that unfolds over centuries or even millennia. In contrast, the North American intellectual tradition holds that progress is an arrow that flies straight and true, unencumbered by conflicts over values or morals. This freedom from the past allows North Americans to focus more on the future. According to Hegel, "America is therefore the land of the future." Einstein expressed similar sentiments: "The American lives even more for his goals, for the future, than the European. Life for him is always becoming, never being." (I found both these quotes in numerous compendia, but was unable to find the original source for either.)

[5] Post-modernism certainly has something important to say about social science in general, including research on violence risk. But the point I am trying to make here is that there are major problems with the predictionist perspective, even when considered within the framework of empirical, positivist social science.
Prediction is neither necessary nor sufficient for making sound decisions about public policy (Brunner, 2000; Rayner, 2000). In the words of Sarewitz, Pielke, and Byerly (2000), “[t]he idea that predictive science can simplify the decision-making process by creating a clearer picture of the future is deeply appealing in principle, but deeply problematic in practice” (p. 4). They point out that many good decisions can be made in the absence of accurate predictions. For example, earthquakes cannot be predicted with any accuracy, but nonetheless one can take many steps to minimize the harmful consequences of earthquakes – improving building codes, distributing survival kits, training children to duck and cover, and so forth. Contrariwise, many bad decisions can be made with the benefit of accurate predictions. For example, predictions of incarceration rates based on race have been used in the United States to justify such things as racial profiling in law enforcement, or even access to voluntary abortion.

Prediction is also misleading. It is often assumed, naively, that prediction is an objective and value-neutral exercise, but this ignores the socio-political context of all human prediction processes and applications. Predictions should not ignore issues related to fairness or legality. Once again, quoting Sarewitz and colleagues, “the idea of a prediction as a disembodied number modified by an uncertainty is entirely too abstract to have any meaning in the real world”; they argue cogently that such numbers are “inextricably enmeshed” in social and political processes (Sarewitz et al., 2000, p. 7).

[6] Many North American scientists who study violence risk deliberately eschew discussion of ethics, morals, values, and politics, attempting to shield themselves from such considerations with the mantle of scientific objectivity. As Jamieson (2000) put it: [W]e Americans tend to be allergic to the open discussion of fundamental moral and political differences.... Pretending that our differences can be washed away in the solvent of scientific decision-making, we instead conduct these arguments in the language of decision theory, or economics, or whatever technical discourse is currently in fashion. This conflation of scientific and policy discourse, foreshadowed in the enthusiasm for science characteristic of founding fathers such as Franklin and Jefferson, has been extremely powerful in American life.... (Jamieson, 2000, p. 321)
But it is impossible to conduct meaningful research or engage in meaningful debate concerning matters of public policy without explicit consideration of values. How can one evaluate without values? Be patient; I have not broken my promise – this perspective is not post-modern, but rather pre-modern. Indeed, it is the Aristotelian notion of phronesis, typically defined as prudence or practical wisdom. The phronetic perspective on the centrality of values in scientific research is discussed clearly by people such as Flyvbjerg (2001) and Jamieson (2000).
Prediction is also distracting. It discourages the active reduction of uncertainty. This is clear in the following quote from Krippendorf (1996, p. 311): “Stochastic population forecasts are a constructive way of incorporating unavoidable uncertainty when making projections of the future.” Efforts to increase predictive accuracy may also divert time and effort away from efforts to prevent or intervene. As I have pointed out previously, I would be happy forensic psychologist if my predictions of violence had near-zero accuracy, as long as this was because there was a near-zero base rate of violence among the high-risk offenders I assessed and treated (Hart, 1998).

2.2 | PREDICTION IS NOT CENTRAL TO THE VALIDATION OF SCIENTIFIC THEORY

Prediction is a weak form of understanding. For example, one can predict that the sun will rise in the east tomorrow without knowing anything about cosmology, astronomy, geography, and so forth. As the philosopher Max Born put it, “You can predict (with the help of a railway timetable) the arrival at King’s Cross of the ten o’clock from Waverly; but you can hardly say that the timetable reveals a cause for the effect” (cited in Cole, 1999, p. 193).

Prediction is more relevant to understanding events rather than processes (Sarewitz & Pielke, 2000). Events are phenomena that can be conceptualized as discrete events and that typically have short or fixed characteristic times, occur in closed systems, and reflect simple or static interactions among elements of the system. In contrast, processes are phenomena that can be conceptualized as multiple events and that typically have long or variable characteristic times, occur in open systems, and are the result of complex, dynamic interactions among elements of the system. When studying processes, accurate predictions about specific events may be impossible or even nonsensical, despite sound knowledge of the process itself. Consider gardening: Even people with the greenest thumbs cannot predict the rate at which particular flowers will do best in a particular bed next season, because many factors critical to floral success – like the weather in the intervening period, or fluctuations in pest populations – are unforeseeable and uncontrollable.

2.3 | PREDICTION IGNORES UNCERTAINTY

Prediction ignores at least four major sources of uncertainty. The first is linguistic uncertainty. As Wittgenstein (1969) noted, language is responsible for many of the problems that plague scientific understanding in the social sciences. Something that cannot be conceptualized and defined clearly cannot be well understood – or well predicted. For example, our failure to establish a consensual, precise definition of “violence” limits our ability to predict whether a person will commit violence.
The second source of uncertainty is evaluative. It is impossible to measure the state of nature precisely due to the principles of relativity, complementarity, and uncertainty (see, for example, Cambel, 1993; Cole, 1999; Press, 1989). These principles come from the big guys: Einstein, Bohr, and Heisenberg. They are believed to limit measurement of physical phenomena, but they are no less relevant to psychological or social phenomena. According to the principle of relativity, measurements can be made from multiple contexts, each of which influences the measurements themselves; but no single perspective or context may be viewed as more correct than the others. According to the principle of complementarity, phenomena may have multiple natures depending on how they are measured; and the very act of measurement may itself alter the nature of the phenomena. According to the principle of uncertainty, there is a fundamental constraint on our ability to measure the initial conditions of an entity, for increased precision of measurement regarding one facet decreases precision of measurement regarding others.

The third source of uncertainty is epistemic. There are, quite obviously, limits to our ability as human beings to perceive and understand the world around us. The work on cognitive biases and heuristics by such people as Kahneman and Tversky and their colleagues (e.g., Kahneman, Slovic, & Tversky, 1982; Kahneinan & Tversky, 2000) and Gigerenzer and his colleagues (e.g., Gigerenzer, 2000; Gigerenzer, Todd, & ABC Research Group, 2000) suggests that we are cognitive misers who process information using a number of highly efficient, although occasionally grossly misleading, heuristics or "short-cuts." Thus, even when human beings have detailed, accurate, complete information available to us, we are unable to make full and coherent use of it. It is also the case, according to Gödel's principle of indeterminacy, that any logical system contains propositions neither provable nor disprovable from its axioms (see, for example, Chaitin, 1982). Thus, some faith is required, even in science.

The fourth source of uncertainty is aleatory or probabilistic. Of course, probabilities by their very nature are uncertain, but it is commonly assumed that they are predictably unpredictable – that is, probability restricts uncertainty within certain parameters. This assumption is untrue. For one thing, probability is a fuzzy concept, with multiple technical definitions (Shafer, 1993). Many definitions of probability stress that it is inherently subjective, something that exists in the eye of the beholder rather than in nature. For example, it is simply impossible to determine objective – that is, frequentist or long run – probabilities for events that have never occurred. Also, it is impossible to calculate some probabilities. Gill, Turing, von Foerster, Chaitin, and others have demonstrated the problems caused by algorithmic or computational complexity (e.g., Chaitin, 1987; see also

[7] Congratulations to Daniel Kahneman, who was awarded the Nobel prize for economics while this chapter was being prepared.
Krippendorf, 1996). For example, relatively simple machines (theoretical or mechanical) can be constructed that defy conventional, “bottom-up” mathematical analysis – there are more potential solutions or structures to the machines than there are atoms on the earth. This has led some to conclude that the analysis of highly complex data may require a “transcomputational” or “top-down” approach, in which analytic decisions are guided by preferences, assumptions, and theories. Finally, recent developments in the mathematics of complexity have made clear that the behavior of even relatively simple deterministic systems can be extremely nonlinear, dynamic, and sensitive to initial values. Chaos theory, catastrophe theory, and their kin point out that minute changes in initial values or imprecisions in measurement can lead to major prediction errors (e.g., Arnold, 1986; Brown, 1995; Gleick, 1987).

The simple, unavoidable conclusion to be drawn from this little review is that it is impossible to make precise quantitative predictions regarding whether a specific person – a sentient, self-organizing, reflective organism – will commit an act of violence at some time in the future.

Now, dear reader, if you are anything like the others who are familiar with this little tirade, there are probably three big questions percolating in your mind right about now. The first big question is this: If predictionism is as bad as I make it out to be, why does our field continue to embrace it? The answer comes, I think, from Gigerenzer’s “tools-to-theories” heuristic: Our acceptance of predictionism is a function of our use of actuarial tools, or statistical prediction methods, in research on violence risk – ironically, the very things to which I would attribute many of our recent advances. I refer to analytic methods such as bivariate and multivariate correlational analysis, logistic regression analysis, Receiver Operating Characteristic analysis, and event history (survival) analysis. Of course, that it serves one well is an excellent reason to use a tool. But another reason to use a tool is that one has no real alternative; that all one has is a hammer. This is a bad reason to use a tool.

The second big question is this: If predictionism and actuarial tools are responsible for the important advances I referred to earlier in this talk, why am I so negative about them? The answer is: The philosophy and tools are too simple to be true. To paraphrase Einstein, our models of the world should be as simple as possible, but not simpler. Scientists must be cautious lest we over-apply the principle of parsimony; we don’t want to end up with a bad case of Occam’s razor burn. For example, correlational analysis, logistic regression analysis, Receiver Operating Characteristic analysis, and event history (survival) analysis – all of these tools, and the predictionist view of the universe they reflect, assume that the outcome is a simple binary or continuous variable that can be evaluated with perfect accuracy. None was designed to deal with an outcome like violence, which is clearly multidimensional and whose measurement is influenced by various forms of bias or error. Also, most of the tools ignore time. They were intended to account contemporaneously for variance in one measure using another, not to
Victims and offenders

make "predictions" about dynamic organisms living in dynamic environments. And none of the tools was intended to handle problems such as violence risk assessment, in which risk factors and the interactions among them grow, change, or evolve over time; and in which the outcome, violence, may not be a discrete occurrence but rather a transactional process that itself changes or evolves over time. The major problem with way-too-simple models and tools is that human beings are basically lazy creatures: We focus on what the models and tools can do; we over-use them, ignore or become complacent with their limitations. We start to think, for example, that maybe a hammer isn’t good for everything, but it’s good enough for most things; maybe the universe isn’t made entirely of nails, but there are enough nails that it’s a waste of time to sweat over the other stuff; maybe we don’t need any new tools, anyway. Or, more on target, we start to think that maybe we can’t predict violence perfectly using simple prediction models, but our predictions are good enough; we can safely dismiss prediction errors, both random and systematic, as unavoidable; and we can start to theorize and communicate about and act on violence risk in terms of numbers and equations without considering alternatives.

The third big question is this: Even if I am right about the problems with predictionism and actuarial tools, what's the alternative? Fortunately, the answer here is simple: Use a new tool. I am of the opinion that further progress in the field of violence risk assessment awaits the development of new tools, both procedural and methodological. Don’t give up the old tools, mind; just try using new ones, as well. Use an approach to understanding, assessing, communicating about, and managing violence risk that reflects a perspective other than predictionism. The tools-to-theories heuristic tells us that changes in the analytic methods we use may cause, rather than be caused by, changes in how we think about things. New tools may lead to new theories. One possible corollary of the tools-to-theories heuristic is that radical changes in theory may flow from radically different tools, and so I would like to suggest that we try a tool that relies as little as possible on predictions, numbers, and equations; one well suited to the description of complexity, change, and transaction; one that almost anyone can be trained to use. The remainder of my talk will be devoted to outlining such a tool, which I will refer to as anchored narrative analysis.

[8] I have a little claw hammer at home, of which I am very fond and make occasional good use. It is great for putting nails in the wall when I want to hang pictures, but not very good for fixing the toilet or cooking.
ANCHORED NARRATIVES

One alternative to the predictionist, actuarial, quantitative approach to risk assessment is an anchored narrative approach. A narrative is simply a story, a communicative device in which multiple incidents are unified into a single entity by means of a plot (Polkinghorne, 1995). The wonderful thing about narratives is that we are all so familiar with them: From the time we are born to the time we die, we listen to stories, read them, and tell them to each other.

3.1 | A LITTLE STORY

The potential benefits of the narrative approach have become clear to me as a result of my participation in various fatality reviews. The first I worked on concerned the sexual homicide of a 12-year old girl, Sarah Kelly, in the small town of La Pas, in northern Manitoba. This was the first time I fully appreciated how crucial stories are in forensic decision-making. The fatality review was, in essence, an attempt to construct a narrative that culminated in the homicide of Sarah. There were several chapters in the story: One focused on the background of the offender; another, on the background of the victim; another, on the law enforcement, corrections, and social welfare services in the community where the offense occurred; one, on the murder itself; and a final chapter focused on the aftermath of the offense. An incredible amount of information was woven together, given meaning and gave meaning, through a narrative analysis conducted by the judge who presided over the review. Since then I have been involved in numerous fatality reviews, a wrongful death suits, and murder trials, as well as more mundane things such as sentencing and workplace violence risk evaluations, and these experiences have served only to strengthen my initial impressions.

I can think of no better way to illustrate what I believe to be the limitations of actuarial tools discussed in the first half of this talk and the strengths of the anchored narrative approach I will describe in the second half than to tell a little story. Based on a fatality review in which I was recently involved, it is the story of a man who, at the age of 23, murdered his mother-in-law. He met a woman at the age of 18, married her, and began to assault her physically. The woman became pregnant by him, and the abuse started to escalate—he was convicted of aggravated assault against her. He left her several times over the next few years, but returned to his wife and to his pattern of domestic violence. The last time he left, the woman decided she would no longer accept him or his violence against her

[9] In discussing the importance of narratives in the study of violence, I am following the lead of others rather than blazing a new trail. Mulvey and Lidz (1985) and, more recently, Litwack (2002) have pleaded for more descriptive research on violence and have conducted such research of their own (e.g., Lidz, Mulvey, Apperson, Evanczuk, & Shea, 1992; Litwack, 1996).
and refused to let him move back into the family home. The man started to stalk
her, even broke into her residence when she was out shopping or visiting friends.
The woman, afraid for her safety, asked her mother to stay with her for a time.
The mother went everywhere with her daughter, answered the phones for her,
nailed the window shut. One night, the man broke into the woman’s house as the
occupants slept. He cut the telephone wires leading into the house, quietly pried
open a window, took off his cowboy boots, climbed into the bedroom, and killed
his mother-in-law by stabbing her 7 times in the chest while she was asleep. He
then went into the bedroom of his former wife, beat and threatened her, held her
and the child captive at knifepoint for several hours, and then was observed by
the police attempting to transport them to another location. He fled, but was cap-
tured a short time later by the police. After his conviction on a charge of murder
in the second degree, the man was sentenced to life in prison with ten years befo-
re parole eligibility. In prison, the man underwent risk assessment conducted by
well-trained and competent professionals. They summarized his risk using state-
of-the-art assessment procedures. Based on these procedures, I would like to
share with you the story about the risk of future violence posed by this man, con-
structed according to predictionist, actuarial principles.

“Bin 3: Low risk; 17% to 30% chance of violent re-offense within 10 years.”

That’s it. Not much of a story, is it? It is the man’s score on the Violence Risk
Appraisal Guide (Quinsey, Harris, Rice, Cormier, & Quinsey, 1998), which is,
arguably, the best means available at the present time of assessing violence risk
using the predictionist-actuarial approach. But the VRAG story doesn’t tell you
about the man’s history of violence, what he has done, or the risk factors present
in the case. It doesn’t tell you what the man might do in the future, or when, or
where, or to whom, or why. It doesn’t tell you what could or should be done to
manage the man’s violence risk. Nor does it provide you with information that
would allow you to figure these things out for yourself. Finally, it doesn’t help you
develop a more general understanding of the causes or nature of violence.

The parole board that saw the man after he served ten year of his sentence was
presented with the full details of this case, including the VRAG findings. The
board, swayed in part by the actuarial risk assessment, released the man. I beca-
me involved with the case after the man was charged with murdering his new
wife – the woman who he met only a few weeks after release to a half-way house;
who shortly thereafter became pregnant by him; who was the victim of physical
assaults perpetrated by him that escalated in severity after the birth of their child;
and who eventually went missing, leaving behind in her residence only residue
of massive quantities of her blood.

I am willing to bet that you learned more from the brief overview of the case –
the little narrative of case facts – than you did from results of the actuarial risk
assessment.
3.2 | A NEW ANALYTIC TOOL

Previously, I stated that a narrative is “simply” a story, but in actuality there is nothing simple at all about stories. Human beings are social, communicative creatures, and narratives are quintessentially human – perhaps the most intricate form of communication or expression known to us. According to Polkinghorne, narrative is “uniquely suited for displaying human existence as situated action…. [It] is the linguistic form that preserves the complexity of human action with its interrelationships of temporal sequence, human motivation, chance happenings, and changing interpersonal and environmental contexts” (1995, pp. 5-7). Narratives are the fundamental means by which we come to understand ourselves and each other. Indeed, it has been suggested that we define our very selves in terms of stories (e.g., MacAdams, 1995).

In mainstream cognitive psychology, as well as many other fields, it has long been recognized that human beings process information very efficiently in the form of narratives. Indeed, Polkinghorne, following Bruner (1985), argued that narrative cognition is one of two fundamental modes of cognitive functioning, the other being paradigmatic cognition – the traditional scientific, deductive, nomological approach to understanding which focuses on “classifying a particular instance as belonging to a category or concept” (Polkinghorne, 1995, p. 9). Paradigmatic cognition is clearly well suited to quantitative analysis, especially of physical objects. Narrative cognition, in contrast, “configures the diverse elements of a particular action in a unified whole in which each element is connected to the central purpose of the action” (p. 11). Narrative cognition defies quantification; it is an inherently qualitative approach to understanding the world, and is especially suited to the analysis of human behavior.

Narrative cognition has its own characteristic rules and operations. For example, a “good” narrative has at least two important features. First, it contains critical elements of information that help to make clear the emotional and motivational meaning of behavior. Second, it has a plot. Plots structure a narrative by establishing its temporal parameters (the beginning and end), guiding the selection of information for inclusion, guiding the order in which information is presented, and clarifying the meaning of information (Polkinghorne, 1995). One can think of information elements as “anchors” for the plot - hence, “anchored narrative.” Astute readers will recognize that the term “anchored narrative” has been used before in psychology, including by Peter van Koppen, Hans Crombag, and colleagues with respect to forensic decision-making (e.g., van Koppen, 1995; Wagenaar, van Koppen, & Crombag, 1993).10

10 Our Dutch colleagues have argued that judges and juries are not Bayesian calculators: They do not consider and weigh information or make decisions in an algorithmic manner. Instead, they actively construct and evaluate possible scenarios that fit the evidence presented to them – including, but not limited to, scenarios put forward by the disputing parties. The judgements of triers of fact depend on how well these scenarios fit all the facts, but especially critical facts or “anchors,” as well as the “goodness-of-fit” of these scenarios with respect to personal experience, implicit theories of human behavior, and so forth.
3.3 | HOW ARE WE TO USE ANCHORED NARRATIVES?

I have described the anchored narrative approach as an analytic tool. But how is one to use this tool to study, evaluate, communicate about, and manage violence risk? There are two basic ways: Narrative analysis and analysis of narratives.

Narrative analysis. Narrative analysis is “the process through which the researcher organizes the data elements [that is, events and happenings] into a coherent developmental account.” In short, the story is the analysis. The task is to construct a story: First, to identify key information elements; and second, to develop a plot that accounts for or ties together these elements. The story must include information that can help inferences about motivations for (i.e., goals of) past acts of violence, and must go beyond that to foreshadow how the person may act in response to possible futures. It is important to recognize that the story is at least partly fictional, insofar as no one knows for certain why people acted the way they did in the past or what will happen in the future; these are merely educated guesses. But if the story is conveyed well – if it includes important information, excludes irrelevant information, supports plausible inferences about motivations, speculates about future conditions in a reasonable manner – then readers will perceive the story to possess verisimilitude and act on it accordingly. If, on the other hand, the story lacks important detail, includes irrelevant detail, contains weak inferences about motivation, or speculates too wildly about the future, it will be – and, I think, should be – disregarded.

In summary, there are at least three general criteria for judging the quality of an anchored narrative:

1. **Comprehensiveness.** The case information gathered should reflect the individual’s functioning in multiple domains and across time, should take into account elements that others consider relevant, and should exclude elements that others consider irrelevant.

2. **Credibility.** The case information should have been gathered from multiple sources and using multiple methods to allow others to evaluate explicitly its consistency.

3. **Clarity.** The plot should have a clear (albeit not necessarily linear) temporal structure, should make clear to others the narrator’s opinions regarding the individual’s motivations and emotions (or allow readers to reach their own opinions), and should assist the development of clear and useful action plans.

Narrative analysis, as I have outlined it here, is a form of qualitative inquiry. Because narratives are not quantitative, it is impossible to evaluate their adequacy with the traditional psychometric methods. Each narrative should be judged on its own merits, as narratives are intended to capture and express the unique features of cases rather than similarities (Flyvbjerg, 2001; Watt et al., 2002). It is possible, however, to evaluate a narrative with respect to criteria such as comprehensiveness, credibility, and clarity.
Analysis of narratives. Once a set of narratives has been generated, it is possible to perform quantitative or paradigmatic analysis of narratives. Here the story is the source of data, rather than the result of the analysis. The researcher develops rules for coding specific elements of the narratives, such as the presence of specific risk factors, developmental trajectories, transactions, or motivational themes. It is possible to evaluate quantitatively the interrater reliability of the coding scheme. It is then possible to examine the frequency with which these elements occur or co-occur in the set of narratives. Analysis of narratives must follow narrative analysis; it is impossible to develop meaningful schemes for coding or classifying narratives until one is familiar with the narratives themselves. Note that analysis of narratives may greatly assist the planning of subsequent large-scale quantitative research; anchored narrative methods can be used to develop new theories, hypotheses, coding schemes, or measure that can be evaluated using the traditional confirmatory methods.

Thus, quantitative methods can be used to evaluate qualitative analyses. In fact, such a procedure has been used several times in the past to judge the quality of forensic reports, most recently in Sweden by Grann and Pallvik to evaluate risk assessments contained in forensic psychiatric reports (Grann & Pallvik, 2002).

3.4 | PRACTICAL EXAMPLES

Let me provide two brief examples of narrative research conducted at the British Columbia Institute Against Family Violence. Both studies focused on intimate partner femicide, which can be defined as the culpable (i.e., non-negligent) homicide of an adult woman caused by a person who was her current or former intimate partner. The first case is a narrative analysis, the second is an analysis of narratives.

Narrative analysis: The Fatality Review Project. The goal of this project (Watt et al., 2002; Watt, Kropp, & Bain, 2003) was to make recommendations concerning the establishment of a fatality review committee in the province of British Columbia, and in particular to determine whether existing sources of information were sufficient to permit a comprehensive review. A multiple case study approach was used to analyze 13 femicides that occurred in the calendar year 1997 in British Columbia. The year 1997 was chosen because it was the most recent one in which all femicide cases were likely to be closed by police. To identify the cases, a list of 30 cases that involved the death of a female aged 18 or older was obtained from the British Columbia Coroner’s Office. Review of coroner’s files and discussion with originating police agencies confirmed that 17 of the cases were intimate partner femicides. For practical purposes, permission was obtained from the two largest law enforcement agencies in the province to review the cases they had investigated, resulting in a final sample of 13 femicides.
Victims and offenders

Information regarding the cases was obtained from standard documentation, specifically, coroner’s reports, police investigations, and newspaper articles. In addition to being the most common sources of information examined by fatality reviews, these documents are quite comprehensive, often including information from autopsy reports, correctional files, health care and social services, interviews with friends and family members, and occupational reports. Information was collected from the files with the use of a comprehensive coding form developed based on a review of the empirical literature, consultation with practitioners, and coding forms developed by previous femicide fatality review committees.

Narrative analysis of the case information was assisted by the development of a standardized script or template. The script was organized both temporally and thematically. The first section identified the victim and perpetrator. The second section summarized the personal histories of the victim and perpetrator prior to the establishment of their intimate relationship, including their family background and child-rearing experiences, school and work adjustment, physical and mental health, past intimate relationships (including attitudes toward relationships), and past criminal and violent behavior (including past intimate partner violence). The third section summarized the history of the intimate relationship between victim and perpetrator, including past intimate partner violence and any contacts with law enforcement, health, social service, or other community agencies. The fourth section described the events immediately leading up to the femicide, including the cause and manner of death.

In many respects, writing narratives in this way mirrors the process of conducting a comprehensive violence risk assessment (i.e., using anamnestic or structured professional judgement procedures) or writing a psycholegal report. Having assisted in the data collection for this project, however, it amazes me that literally thousands of bits of information concerning a case – the police statements of perpetrators, the personal journals of victims and perpetrators, interviews with family and friends, crime scene photographs, autopsy reports, newspaper articles – can be synthesized in the form of a relatively short story, typically about 5,000 to 10,000 words in length. And once the information was presented in narrative form, it could be analyzed intuitively; in fact, it is impossible as a human being not to analyze such a story. Some of the narratives made perfect sense: Patterns of conflict or violence evident in past relationships started to emerge in the new relationship; and the perpetrator’s behavior in the days or weeks before the murder revealed a clear motive for the femicide, and the evolution of motive into plan and, finally, action. Other narratives made no sense at all; the story, or the attempt to construct a story, made obvious that critical information was missing or that the motive may have been unusual.

Although this project is still underway, several important findings already have emerged. One is that police files, coroner’s reports, and newspaper reports did not provide systematic information about the roles played by various community agencies or about the backgrounds of victims and perpetrators. This was espe-
cially true when a thorough police investigation was unnecessary because the perpetrator confessed or died by suicide after the femicide. Another is that most schemes developed for coding risk factors were woefully inadequate, assuming that risk factors are simple and stable. A perfect example is something as basic as “marital status.” A victim in this project had multiple common-law relationships in the past. At the time of her death, she was co-habiting with four different partners on different nights of the week – all of whom knew about each other – and talking to a fifth man about entering into a relationship with him. The man who eventually killed her was a former common-law partner who was also renting a room in one of the victim’s residences. It is difficult to capture this woman’s marital status, or the victim-perpetrator relationship, in a single category. A third example is that, contrary to stereotypical notions, very few of the femicides were “crimes of passion.” In 12 of 13 cases, there was a period of planning or preparation, including such things as explicit threats or expressed homicidal intent or the purchase of weapons, that preceded the femicide by days or even weeks.

Analysis of narratives: The Conditional Release Project. The goal of this project (Hart, Kropp, & Bain, 2002) was to review apparent cases of serious family violence committed by federal offenders on conditional release in Canada during the past 10 years. We were asked to identify potential problems in offender management before and after conditional release that might assist the revision of policies and procedures concerning conditional release that might prevent such occurrences in the future. Of the 10 reports reviewed, 9 concerned apparent femicides; the other involved kidnapping and sexual assault. I will discuss only the femicide cases here.

An analysis of the narrative reports issued by boards in investigation in the cases was conducted. The original investigations were operational reviews intended to identify any potential issues related to non-compliance with policies and procedures. The various boards had inspected relevant documentation available from all correctional, parole, and law enforcement agencies involved in the case, and also conducted interviews with agency personnel. The reports all had a similar structure, beginning with a description of the (alleged) critical incident and followed by a summary of the offender’s criminal history, a chronology of events, an analysis of key issues, and a summary of key findings. Most of the reports ended by making recommendations for improving service delivery based on the issues and findings of the case at hand. The reports ranged in length from about 10,000 to about 30,000 words.

The narratives in this project already had been constructed, so we reviewed them to identify thematic similarities with respect to the issues, findings, and recommendations outlined in the reports. Given the small sample size, we did not conduct quantitative analyses but instead supported our findings by listing the cases in which particular themes were apparent. All data came from these reports; no attempt was made to gather new information.
We made a number of conclusions and recommendations based on our analysis of the narratives. Some reflected themes that were present or addressed in the reports. For example, we noted that all offenders had a history of serious violence prior to conditional release; in most cases (8 of 9), this included an officially documented history of family violence. Yet, the nature of offenders’ violence history was reported by agency staff in a manner that was often superficial, inconsistent, or inaccurate. On the basis of this and similar conclusions, we recommended that agency staff be provided with training and assessment instruments for the investigation of offenders’ history of violence – and especially their history of family violence. As another example, we noted most of the offenders entered into new intimate relationships during conditional release – sometimes several new relationships. Often, the new spouse entered the relationship with children or the offender and his new spouse had children during conditional release. But rarely were new spouses the focus of systematic evaluation, and the new spouse or children were not identified as potential victims of family violence; in none of the cases was victim safety planning initiated. On the basis of these conclusions, we recommended that agency staff should engage in safety planning with (potential) victims, including (ex-) spouses and children, prior to the conditional release of offenders with a history of serious family violence.

Other conclusions and recommendations we made reflected themes that were absent or not addressed consistently in the reports. For example, we noted that a few of the reports contained sophisticated treatment of family violence issues, but most did not. We interpreted this to indicate that considerable expertise in the field of family violence existed within the agencies, but was not distributed equally across their operational units. We therefore recommended that the agencies consider concentrating family violence expertise in specialized units that would consult with staff across the nation.

3.5 | ADVANTAGES OF THE ANCHORED NARRATIVE APPROACH

As I see it, there are five major advantages to conceptualizing violence risk assessment within an anchored narrative approach. First, as discussed already, the narrative approach is ideally suited to the analysis of human behavior, especially in psycholegal contexts. It focuses on inferring from key facts the motivations, emotions, and cognitions of actors – factors that typically are ignored in predictionist-actuarial approaches. The legal system assumes, of course, that people are thinking, feeling, decision-making creatures, so it is hardly surprising the few existing studies that evaluated various types of reports found lawyers and judges preferred narrative, idiographic, descriptive reports to quantitative, nomothetic, actuarial reports.

Second, the narrative approach attempts to understand the influence of context on actors. In North America, people such as Hank Steadman, Ed Mulvey, and
VIOLENCE RISK ASSESSMENT

(more recently) Eric Silver have been strong proponents of including certain contextual or situational factors in our models of violence. But at present it is simply impossible to do good quantitative research on context: We don’t have a good definition of context (social, psychological, geographic?) or framework for conceptualizing it (facets, layers, interactions?), let alone a good way of measuring it. The narrative approach may go a long way towards helping else understand, define, and measure the concept of context.

Third, the narrative approach is ideally suited to the analysis of developmental change in people, contexts, and relationships (e.g., Watt, Allen, Kropp, & Bain, 2002). In the field of violence risk assessment there is a myth, widely held and long perpetuated, that there exist important “static” risk factors. But virtually all these factors change over time, or at least some facet of them changes, or their causal relevance changes. For example, one’s birth date doesn’t change, but one’s age changes continuously. Chromosomal sex doesn’t change, but sexual characteristics and gender identity can and do change over time, sometimes radically. Criminological variables such as “prior record” or “age at index offense” change with each new conviction. Symptoms of all mental disorders fluctuate over time in severity. Why do we code these risk factors using simplistic schemes and then blame the risk factor for being useless when we can’t find evidence of strong, clear, direct influences on violence? A narrative approach should help us to appreciate the wonderful complexity of these factors, to develop new and better coding schemes for future quantitative research.

Fourth, the narrative approach is ideally suited to the analysis of transactions among risk factors (Watt et al., 2002). Not only are most risk factors dynamic over time, but furthermore they change each other, they exert mutual or bi-directional influences on each other – and the nature of the transaction may itself change over time. For example, think about the association between substance use and major mental illness. There is some evidence that mental illness potentiates substance use as a risk factor for violence; the two factors interact in a multiplicative or synergistic manner. I would also wager that factors such as personality disorder or neuropsychological abnormality may further complicate the interaction, and that the synergy between substance use and major mental illness may be strongest in certain contexts (for example, when the individual is living in the community with poor access to treatment and social services) and at certain developmental stages (for example, when the individual is young and the disease process is in its early or formative stages). As another example, consider the relevance of the victim-offender relationship in domestic violence, stalking, sexual assault, or child abuse. Transactions between the victim and offender can have a profound influence on violence, in some cases leading to a dramatic escalation in frequency or severity, and in other cases leading to the desistence of violence.

[11] Here, following Watt et al. (2002), I use the term transaction to refer to a specific type of interaction in which risk factors change each other as well as the outcome of interest.
Fifth, the narrative approach is ideally suited to interpersonal communication, both among professionals and between professionals and consumers or decision makers. All psychologists know that numbers are difficult to understand and discuss—even with people who work with numbers for a living. The construction of narratives is, on the hand, a highly intuitive and familiar process. In fact, narratives are already the stock-and-trade of communication in professional psychology, as well as in the legal system. Furthermore, whereas the practical relevance of numbers may be unclear, the implications of a narrative generally are much clearer. In fact, one can conceptualize a good violence risk assessment as a contingency-based action plan for what should be done in the future, rather than a quantitative statement of fact about the future (Hart, in press). Because it is a plan concerning a human being, to be read and implemented and evaluated by other human beings, a risk assessment should be linguistic rather than quantitative in nature.

3.6 | OBJECTIONS TO THE NARRATIVE APPROACH

When I have conversed with people about the anchored narrative approach to violence risk assessment, some have voiced concerns. Most of them reflect a fear of “methodological anarchy” — a fear that using a non-actuarial tool will lead to an intellectual Dark Age in which all the achievements of predictionism will be lost to us. There are three major variations on this theme:

1. The narrative approach is fuzzy and subjective. There is concern that no rules govern narratives, and therefore there can be no consistency in the stories that people construct about violence risk. In the absence of objective criteria, evaluators will be free to decide for themselves which information is relevant and should be considered in risk assessment, and consumers will have to rely on the persuasive writing of evaluators to determine the credibility of assessments.

My response is that although the narrative approach is indeed subjective, there certainly are rules for judging the quality of a story, and most people know these rules intuitively. Although at present the rules are admittedly rather general and implicit, I see no reason why they cannot be made more detailed and explicit.

Also, most people who raise this argument have false impressions regarding the objectivity and precision of actuarial tools. Anyone familiar with actuarial tools knows that their construction is anything but “objective.” There are a multitude of decisions instrument developers must make, including such things as which variables to analyze, how to code them, or which statistical procedures to use for selecting and weighting them. This means that developers are free to construct their tools as they see fit—often selecting a half dozen or dozen of their favorite variables, coded in very specific ways, from a pool of many thousands. Also, I have had considerable opportunity to review the oral and written testimony of clinical-forensic psychologists in legal proceedings, and the consensus or interrater reliability for opinions based on actuarial analysis of cases is no better than that
for opinions based on narrative analysis. Furthermore, most consumers are completely unable to judge the validity of actuarial procedures, and so must rely on the (subjective) opinions of the experts who develop or use such procedures. So how is the actuarial approach any better than the narrative approach? Actuarial tools are pseudo-objective and pseudo-precise; at least narratives make no pretense at either objectivity or precision.

2. The narrative approach is not empirical. Another concern is that narrative analysis, like every qualitative (i.e., non-quantitative) method, is inherently unscientific — perhaps even anti-scientific. My response is that the hallmark of modern science is empiricism, which involves observation as much as experimentation; and furthermore, that quantification is neither necessary nor sufficient for good observation or experimentation. Any good scientist makes careful, systematic observations about the phenomena she investigates; sound experimental and statistical analysis, if necessary at all, cannot take place without the benefit of observational research. The summary dismissal of qualitative inquiry reflects methodological naivety or, worse, chauvinism.

3. The narrative approach is untestable. The final concern is that narratives cannot be evaluated quantitatively. My response is that this is partly true: It is indeed difficult to reduce stories to numbers that can be crunched on your laptop using statistical software. But that’s precisely why the narrative approach has the potential to provide new insights! Also, I have discussed already how it is possible to use quantitative methods to evaluate at least some aspects of the narrative approach.

**4 | CONCLUSION**

We cannot afford to delude ourselves that we can predict future violence in a given case with any reasonable degree of scientific or professional certainty. If we do, we will stop striving for a deep understanding of nature of violence and allow ourselves to be used as justification for draconian political decisions and social policies (Hart, 2002). It is time that forensic psychologists outgrew the predictionist philosophy of our youth and started to do our jobs differently, to add a tool to our toolboxes. It is time for us to begin writing a new chapter in the story of our field.

**REFERENCES**


244 Victims and offenders


Violent and non-violent young offenders

A COMPARISON OF DEMOGRAPHICS, MULTI-PROBLEM PROFILES AND PSYCHOPATHIC TRAITS

RAYMOND R. CORRADO, IRWIN M. COHEN & FRANCO MARINO

During the last 20 years of research on violent young offenders, the major accomplishment theoretically has been the identification of a full range of risk and protective factors that are correlated with violent behavior. Most of the identification of these factors has emanated from the cohort and panel studies conducted mainly in the United Kingdom, the United States, Sweden, New Zealand, and Canada. Arguably, the most dramatic recent illustration of this research trend is the recent work by the research team from the MRC Social and Genetic and Developmental Psychiatry Centre at the Institute Psychiatry in London led by Terrie Moffitt. According to this study, while individuals who had a combination of low activity Monoamine Oxidase (MAOA) genotype and maltreatment were only 12% of the male birth cohort, they accounted for 44% of the cohort’s violent convictions (Caspi et al., 2002). Moreover, 85% of the cohort males who had this genetic makeup and were severely maltreated developed some form of antisocial behavior (Caspi et al., 2002). These researchers contend that their data demonstrates the causal link between a person’s genetic make-up and sensitivity to environmental factors in regard to explaining violence. We believe that this research team’s findings are critical since they help understand the enormous complexity of integrating risk factors into a parsimonious theoretical model. One of the challenges in this regard is the often low correlations between well-known risk fac-

[i] This research was supported by grants from the Social Science and Humanities Research Council of Canada (R-410-98-1246 and R-410-2001-1132) to Dr. Raymond R. Corrado.
tors and violent adolescents. For example, not all victims of maltreatment become violent, yet maltreatment appears to be a crucial condition that appears causally activated with a genetically determined low level of MAOA.

Other cohort studies have identified several risk factor trends for violence. In addition, to the original studies, such as the groundbreaking research of West and Farrington and the Cambridge-London project, other cohort sample studies have been conducted in the last two decades by numerous leading scholars (Loeber, Farrington, & Waschbusch, 1998; Hawkins, Laub, & Lauritsen, 1998; Huizinga and Jakob-Chien, 1998; Lipsey and Derzon; Hawkins et al., 1998; Loeber and Farrington, 1998; Offord, Lipman, & Duku, 2001; Keenan, 2001; Tremblay and LeMarquand, 2001; Caspi et al., 2002). While it is not possible to provide even a cursory summation of the enormous amount of research that identifies each risk factor for violence, we believe that it is important to identify certain empirical and theoretical gaps in the current research and how our current research project on a sample of serious and violent incarcerated young offenders in British Columbia can contribute to a more complete understanding of these lacuna. It is important to point out that there are several studies under way involving most of these above-mentioned renowned researchers that will address the current gaps. However, the research that we are currently conducting has confirmed that certain personality traits identified with psychopathy appear to be important in understanding serious and violent young offenders.

Loeber and Hay (1994) outline a three-pathway model to explain serious and violent young offenders. Like Moffitt (1997), Patterson, Capaldi, and Bank (1991), and Thornberry and Krohn (in press), the key distinction in explaining the developmental trajectory of serious violent offenders and serious nonviolent offenders can usually be traced back to individual, family, and socio-economic factors in early childhood. Loeber identifies variables such as authority conflict, stubbornness and defiance, and later childhood and early adolescent aggressive behaviors such as bullying. In one study that tested Loeber’s model, Tolan and Gorman-Smith (1998) found that early age of onset of anti-social behavior is correlated with the three-pathway model. However, the model did not work well with a high-risk inner city sample. Similarly, the model did not apply for females. However, as Loeber has admitted, and Tolan and Gorman-Smith have argued, it may be that a single early onset pathway model might not explain serious versus violent offenders. The latter researchers suggest that there are likely multiple ways to singular acts of violence and violent offenders as there is likely a single pathway to serious offenders.

Moffitt (1997) distinguishes pathways according to early onset for “life course-persistent” and late onset “adolescent-limited” types. The former pathway is characterized by children who exhibit very early neuropsychological deficits including reading and speech problems, and executive function deficits, such as inattention, impulsivity, and aggression. These children pose an immediate challenge to parents with their aggressive, defiant, and disruptive behaviors. Parents who are
ineffective in disciplining, managing, and teaching pro-social behavior, and/or experience family adversities contribute to a reinforcing sequence of “failed parent/child encounters” that cause “persistent anti-social behavior” throughout adolescence and adulthood. This behavior, often violent, persists because of “contemporary continuity”, i.e., early problems set in motion a series of subsequent problems in a chain reaction sequence. This chain consists typically of neuro-cognitive executive function deficit that results in reinforcing ineffective parenting. This, in turn, leads to difficulty for the child in learning to substitute pro-social behavior for antisocial behavior followed by peer and adult rejection. This results in a reduction of educational skills causing difficulty for the youth in successfully making the transition to adult conventional rules and reduced opportunities for the youth to break the sequence causing antisocial behavior. In contrast, adolescent-limited offenders do not have the biological deficits and, therefore, do not develop cumulative disadvantages. They are frustrated by the gap between biological age and social age, and, consequently, want to connect with other adolescents, often late course-persistent youth, where they learn to “socially mimic” delinquent behaviors. However, when late adolescence and early adulthood desirable pro-social options become available, they desist from crime (Moffitt, 1997).

More recently, Moffitt and her colleagues have identified the key difference in pathways between violent and non-violent offenders as the early childhood (3-5 ages) temperament characteristic ‘lack of control’. This temperament consists of “an inability to modulated impulsive expression, impersistance in problem solving, as well as sensitivity to stress and challenges that is expressed in affectively changed negative reactions” (Caspi et al., 1995:59). In effect, fleeting inattention and emotional liability define the temperament characteristics of children who are likely to be violent at 18. In contrast, certain child and family factors, such as the number of changes of primary caregivers, the number of residence changes, and single parent status, did not distinguish between violent and non-violent offenders.

As mentioned above, Caspi et al. (2002) have posited a greater interactive effect with child maltreatment and later violent offenders. This is a critical finding as it provides another pathway to violent offending or an elaboration of a single pathway for life course-persistent offenders who likely constitute a large proportion of violent offenders. As well, Moffitt et al. (2001) specify individual-level traits that best distinguish both boys and girls and within gender differences in antisocial behaviors including violent offenders. These traits are neuro-cognitive deficits, undercontrolled temperament, weak constraints, and hyperactivity. The later three traits constitute key dimensions of psychopathy. As well, Caspi et al. (2002) contend that these traits, especially neuro-cognitive deficits, are considerably more important than the classic family and parental high risk variables in explaining extreme violent behaviors and its preponderance among boys.

However, there are theoretical differences among researchers concerning the primary causal mechanisms. Patterson, Reid, and Dishion (1992) focus on the inability of parents to react appropriately to the coercive processes instigated by early
Victims and offenders

childhood misbehavior. Instead of responding to tantrums and aggressive behaviors with consistent and proportionate discipline, parents provide negative reinforcement by allowing the child to get his or her way. This repeated negative reinforcement leads to early anti-social behavior that sets in motion the sequence of future failed social relations and poor academic experiences followed by rejection from prosocial peers and associations with other anti-social children and adolescents (Patterson, Reid, & Dishion, 1992).

A different perspective is proposed by Thornberry and Krohn (in press). They argue that there are distinctions between ‘early starters’ and ‘later starters’. The ‘early starters’ are distinguished by the high degree of individual risk problems faced by the child and family. These problems occur in a context of structural adversity, such as poverty, racism, and family conflict that together constitutes an accumulated disadvantage. It is this structural adversity that reduces the social capital that parents need in order to cope with the child’s negative temperamental qualities (Thornberry and Krohn, in press). Poor family management skills are linked to poor parenting skills, which, in turn, reinforce the child’s negative individual traits. Structural adversity continues throughout the life-course of the child and the subsequent persistent anti-social behavior causes the continuing negative relationships with parents and pro-social peers and poor academic performance. As children and adolescents, these children become trapped in deviant lifestyles. In contrast, Thornberry and Krohn argue that ‘late starters’ are less likely to experience these extreme and interconnected causal factors of structural adversity and negative temperamental qualities.

Gottfredson and Hirschi’s (1990) self-control theory also focuses on ineffective parenting and the development of low self-control. However, they do not explain the developmental causal mechanisms to the degree that the other theorists discussed above do. Nonetheless, Gottfredson and Hirschi set their individual-level trait in the context of Hirschi’s early theory of social control and bonding. Similarly, Wilson and Hernstein (1985) argue that the primary cause of violence, again, is the inability of certain children and adolescents to defer gratification. According to Wilson and Hernstein, impulsive, non-empathic, and aggressive personality traits cause violence and these traits are independent of career criminal pathways predominant in the above theories.

The more specific mechanisms likely involved in the evolution of negative personality traits or disorders included in these theories have been most elaborated in developmental psychology. The typical sequence begins with ‘difficult’ temperaments and how these temperaments cause low self-control, both of which are linked to poor management of negative emotions. The latter is negatively affected by poor language development, poor attachment to caregivers, poor parenting, and family adversity. The interactive or reciprocal causal effects of all these variables have a cumulative negative effect on the development of empathy. The impressive research in developmental and educational psychology is best exemplified by Richard Tremblay and his colleagues (2000) and the Dunedin project.
in New Zealand lead by Terri Moffitt (2001). Again, the cumulative research can be narrowed to personality and emotional characteristics and their interactive effects with parenting. The variables that are consistently linked to serious and/or violent offenders are high behavioral activation, such as hyperactivity, and low social reward dependence, such as dependence on rewards from pro-social behavior (Tremblay and Marquard, 2001). Low anxiety, short-attention span, emotional lability, and negativism at the childhood stage, and then adolescent developmental disorders, such as conduct disorder, are related constructs that appear consistently in this developmental paradigm.

A more recent research trend in the attempt to specify the needed causal mechanisms for distinguishing violent and non-violent offenders is psychopathy. Among the theorists who identified early psychopathic traits were Eysenck and Eysenck (1976) who utilized the construct of psychotism to describe “cold, egocentric, and tough-minded” youth. However, these traits more typically are part of the traits identified by the construct psychopathy. This personality disorder consists of two dimensions: interpersonal-affect and behavioral traits. The interpersonal traits include superficiality and egocentricity, while the affective traits include shallowness and callousness. The behavioral dimension consists of impulsive and risky acts. Most of the research, including the development of the clinical rating scale the Psychopathy Checklist- Revised (Hare 1991), focused on assessing this disorder in adult criminal offender populations. This research indicates that psychopathy is a reliable and valid construct with considerable predictive validity regarding criminal behavior. Specifically, psychopaths commit the majority of crime (Hare, 1998), are more likely to be violent offenders (Hare, 1998), and have higher rates of recidivism (Hemphill, Hare, & Wong, 1998).

More recently, there is a growing body of research on the assessment of psychopathy among certain new populations, usually incarcerated delinquent samples. Theoretically, it is argued that, given that psychopathy is a personality disorder, its traits would be evident early in a child’s development. As we have demonstrated above, virtually all of the research involving risk factors and theories explaining and predicting violent offenders identify many of the personality traits incorporated in the psychopathy construct. Hare and his colleagues (Hare, Forth, & Strachan, 1992; Hare, 1998) and, particularly, Forth (Forth, 1995; Forth, Hart, & Hare, 1990) have maintained the theoretical validity of utilizing psychopathy for explaining delinquent and/or criminal behavior among children and adolescents. The PCL-R has been reliably and validly utilized in samples of adolescent offenders in a manner similar to adult samples (Brandt, Wallace, Patrich, & Curtin, 1997). Regarding its predictive validity, several studies demonstrate that psychopathy is predictive of violence (Brandt et al., 1997; Toupin, Mercier, Dery, Cote, & Hodgins, 1996) in various settings, including detention centers and inpatient hospitals (Forth, Hart, & Hare, 1990; Myers, Burkert, and Harris, 1995).

This research resulted in the development of the clinical rating scale, the Psychopathy Checklist: Youth Version (PCL-YV) (Forth, Kosson, & Hare, in
press). Like the adult version PCL-R, the PCL-YV also consists of the two afore-
mentioned dimensions of interpersonal-affect and behavioral traits. However,
several items have been adjusted to take into account the more limited life expe-
riences of adolescents, such as the lower likelihood of marital relationships. Still,
with the innovative measurement and theoretical development of psychopathy in
understanding adolescent criminality, there is a growing concern about the
validity and ethicality of assessing a personality disorder originally created for
adults (Edens, Skeem, Cruise, & Caufman, 2001; Vincent and Hart, 2002). The
key ethical issue is that psychopathy has such a powerful negative connotation
that both the public and criminal justice officials may respond far too punitively
to adolescents diagnosed as psychopaths. Nonetheless, there is the counter argu-
ment that the most effective interventions occur when personality disorders are
identified at the earliest developmental stages (Salekin, Rogers, & Machin, 2001).
Moreover, contrary to popular images, there is no consensus among clinical psy-
chologists that psychopathy cannot be effectively treated. Partly in response to
these concerns, considerable research is ongoing to investigate the construct
validity of adolescent psychopathy assessments (Lee, Kolen, Frisbie, & Ankenmann,
2001; Vincent, 2002). In the research findings presented below, we
utilized the PCL-YV (see Figure 1) with a two-factor structure.

**Figure 1. The psychopathy checklist: youth version**

<table>
<thead>
<tr>
<th>Factor 1: Interpersonal/Affect Traits</th>
<th>Factor 2: Behavioral Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Glibness/Superficial Charm</td>
<td>• Need for Stimulation/Proneness to Boredom</td>
</tr>
<tr>
<td>• Grandiose Sense of Self-Worth</td>
<td>• Parasitic Lifestyle</td>
</tr>
<tr>
<td>• Pathological Lying</td>
<td>• Poor Behavioral Controls</td>
</tr>
<tr>
<td>• Conning/Manipulative</td>
<td>• Early Behavioral Problems</td>
</tr>
<tr>
<td>• Lack of Remorse/Guilt</td>
<td>• Lack of Realistic Long-Term Goals</td>
</tr>
<tr>
<td>• Shallow Affect</td>
<td>• Impulsivity</td>
</tr>
<tr>
<td>• Callousness/Lack of Empathy</td>
<td>• Irresponsibility</td>
</tr>
<tr>
<td>• Failure to Accept Responsibility</td>
<td>• Juvenile Delinquency</td>
</tr>
<tr>
<td>for own Actions</td>
<td>• Revocation of Conditional Release</td>
</tr>
</tbody>
</table>

In order to assess psychopathy in a sample of serious and violent incarcerated young offenders, a team of graduate and undergraduate researchers was trained on the PCL-YV and the Psychopathy Checklist – Screening Version (PCL-SV). The information used to assess psychopathy derived from in-depth interviews with consenting youth and a thorough review of their institutional files. In addition, inter-rater reliability checks were routinely performed with all researchers.
The combination of interviewing the offenders and reviewing their files provided a great deal of information on a wide range of issues.

The data presented in this chapter is based on 525 youth who were serving a period of incarceration in one of four custody centres; two closed custody facilities and two open custody facilities. The primary difference between open and closed custody in Canada is the level of security that offenders experience and their ability to access community resources. In closed custody, offenders are granted significantly less freedom of movement during the day, are under increased levels of surveillance at all times, and have less opportunity to access community resources and programs. Of the 525 youth who make up this sample of serious and violent young offenders, 400 are male and 125 are female. With respect to ethnicity, while the majority of the sample is Caucasian, 21.4% of the sample self-identifies as Aboriginal. This represents a significant overrepresentation of Aboriginal youth as the population of Aboriginal people in British Columbia under the age of 18 is estimated to be around 3%. Not only are Aboriginal youth overrepresented in the sample, they are grossly overrepresented in the category of violent offenders. Of those youth who self-identified as Aboriginal, 66% are violent offenders. The age range of the youth in our sample is from 12 to 19 years of age with a mean age of 16 years old.

In terms of violent offending, 69% of the youth in our sample are defined as violent offenders. This classification is based on having been convicted at least once of a violent offence. Of the males in our sample, 74% have been convicted of a violent offence, while 66% of the females have at least one prior conviction for a violent offence. The PCL-YV was completed on 297 youth with 17% of this subsample meeting the threshold for psychopathy by scoring 30 or more on the PCL-YV. Of the 297 youth who have PCL-YV scores and meet the criteria for a designation of psychopathy, 92.2% are violent offenders. Table 1 demonstrates the most serious lifetime offences for the subsample of youth with PCL-YV scores. As expected, psychopaths have engaged in more serious violent offending than their non-psychopathic counterparts.

Table 1. Most serious lifetime offence resulting in a conviction

<table>
<thead>
<tr>
<th>Offence</th>
<th>Psychopaths (n=51)</th>
<th>Non-Psychopaths (n=246)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delinquent Offences</td>
<td>0</td>
<td>0.4%</td>
</tr>
<tr>
<td>Drug Offences</td>
<td>0</td>
<td>1.2%</td>
</tr>
<tr>
<td>Property Offences</td>
<td>7.8%**</td>
<td>24.4%**</td>
</tr>
<tr>
<td>Minor Assaults</td>
<td>33.3%*</td>
<td>24.4%*</td>
</tr>
<tr>
<td>Serious Assaults and Robbery</td>
<td>43.1%</td>
<td>43.9%</td>
</tr>
<tr>
<td>Sexual Offences</td>
<td>7.8%</td>
<td>4.5%</td>
</tr>
<tr>
<td>Murder/Manslaughter</td>
<td>7.8%</td>
<td>1.6%</td>
</tr>
</tbody>
</table>
With respect to the offences for which the youth in our sample are currently serving a period of incarceration in either an open or a closed custody facility, while there are high levels of violent and serious violent offences, a large portion of our sample (57.4%) were in custody for property or delinquent offences. Of the psychopaths in our sample, 56.9% were currently serving a period of incarceration for a non-violent offence. This finding may reflect the process of escalating sentences in which minor offences result in incarceration because of the habitual nature of the offender or their possible prior conviction(s) for violence. Still, the seriousness of this sample’s offending profile is demonstrated in the fact that, without consideration of the youth’s prior offending history, 3.6% of the youth were convicted of murder, 2.5% for sexual offences, 22.3% for robbery, and 30.2% for assaults. For those who were rated a psychopath on the PCL-YV, 7.8% were currently serving a period of incarceration for murder, 25.5% for assaults, and 9.8% for robbery.

The serious nature of our sample is demonstrated by the mean amount of time that these youth have spent on probation and/or in custody during their lives. Comparatively, the non-psychopathic youth in our sample have spent 3 years on probation, 4 months in open custody, and 4 months in closed custody. In other words, psychopaths have served a year more than non-psychopaths on probation and 3 more months in closed custody (p < .001). It is also important to note that the mean number of prior sentences for the non-psychopaths is 3.9, but the number of previous convictions for the psychopaths is 6.3. With respect to the key risk factor of the youth’s age at first contact with the criminal justice system, the mean age is 14.4 for violent offenders, 14.7 for non-violent offenders, and 14 for psychopathic offenders (p < .05).

As mentioned above, predictor variables typically fall into categories related to demographics, familial, and psychological characteristics. In terms of the stability of a youth’s living conditions, 41.1% of the non-violent and 48% of the violent offenders was living with their immediate family (p < 0.001) at the time that they committed the offence for which they are currently incarcerated. For the psychopaths, 37% were living with a member of their immediate family. In most cases where the offender was living with an immediate family member, these families were headed by a single mother. Moreover, the sample is composed of a large number of violent youth (44.1%) who are living on the streets, by themselves, or as wards of the state, and, among the sub sample, a large number of psychopathic youth (45%) who were living independently. While there are moderate levels of parental employment for the entire sample, it must be kept in mind that parental employment may not be a mitigating factor for these youth as many come from broken homes or do not live and/or have any contact with their employed parent(s).

The lives of the youth in our sample are characterized by extremely high levels of family dysfunction. Family dysfunction has frequently been considered linked to risk for violent offending. The vast majority of youth have left their primary home
by their own choice. Interestingly, more non-violent youth (87.4%) than violent youth (76.5%) have decided to leave their primary home (p < .05). Moreover, the number of times that these youth have selected to leave their homes is extremely high. While not statistically significant, it is interesting to note that non-violent offenders have left home more times (15.3 times), on average, than the violent offenders (9.6 times). It is also distressing to see the mean age at which both violent and non-violent youth first decide to leave their homes is 12.8 years old. For the psychopaths in our sample, 85% have left their primary home by their own choice as compared to 75% of their non-psychopathic counterparts. The mean number of times that psychopaths have left their home is 14 times, while non-psychopaths have left their home 10 times. The age that these youth began leaving their homes is 12.6 for the non-psychopaths and 12.2 for the psychopaths.

In addition to the high rates of leaving home by their own volition, 50.5% of the non-violent offenders and 47.5% of the violent offenders have been kicked out of their homes. It is interesting to note that the non-violent offenders have been kicked out of their homes more often than the violent offenders. Specifically, violent offenders have been kicked out of their home 4.3 times while non-violent offenders have been kicked out 7.7 times (p < .05). Both violent and non-violent offenders began being kicked out of their home at 13.4 years old. For our sub sample of those with PCL-YV scores, 56% of the psychopaths and 46% of the non-psychopaths have been kicked out of their homes. Psychopaths have been kicked out of their home 4.1 times, while their non-psychopathic counterparts have been kicked out 5.7 times and the age at which these youth began being kicked out of their home is 13 years old for both samples. Moreover, these youths’ lives of characterized by a high level of mobility. In terms of places other than their primary home in which a youth has stayed at for at least 3 consecutive nights, violent offenders have lived, on average, in 8.2 other places, while our non-violent offenders have lived, on average, in 9 other residences. While not a statistically significant difference, the psychopaths in our sample have lived in 12.29 other places as compared to 8.6 other places for the non-psychopaths.

High levels of family dysfunction are also evident by examining the histories of the family members of these young offenders. As demonstrated in Table 2, there are high levels of alcohol abuse, drug use, a family member having been sexually and/or physically abused, a family member who suffers from a mental disorder or illness, or having a family member with a criminal record. It is important to note that we are not able to temporally link these family problems to the youth’s offending history or the emergence of any psychopathic traits.

There is a statistically significant higher portion of non-violent offenders who report at least one family member as having a drug abuse problem. There are also extremely high rates of family members having a criminal record. Having at least one family member who has been sexually abused is also statistically significant for our non-violent offenders. While family dysfunction is considered a risk factor for violent young offending, in all cases, the non-violent offenders report hig-
Victims and offenders

Table 2. Self-reported family problem profile

<table>
<thead>
<tr>
<th></th>
<th>Non-Violent Offenders (n=143)</th>
<th>Violent Offenders (n=364)</th>
<th>Psychopaths (n=51)</th>
<th>Non-Psychopaths (n=246)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol Use/Abuse</td>
<td>76.4%</td>
<td>72.5%</td>
<td>78%</td>
<td>74.9%</td>
</tr>
<tr>
<td>Drug Use/Abuse</td>
<td>67.4%***</td>
<td>48.4%***</td>
<td>56%</td>
<td>59.1%</td>
</tr>
<tr>
<td>Victim of Physical Abuse</td>
<td>47.2%</td>
<td>49%</td>
<td>54.3%</td>
<td>49.8%</td>
</tr>
<tr>
<td>Victim of Sexual Abuse</td>
<td>26.4%*</td>
<td>17.6%*</td>
<td>30.4%</td>
<td>21.8%</td>
</tr>
<tr>
<td>Mental Illness</td>
<td>24.3%</td>
<td>20.9%</td>
<td>18%</td>
<td>24.6%</td>
</tr>
<tr>
<td>Criminal Record</td>
<td>66.7%</td>
<td>65.4%</td>
<td>74%</td>
<td>66.3%</td>
</tr>
<tr>
<td>Foster Care</td>
<td>44.4%*</td>
<td>31.4%*</td>
<td>41.7%</td>
<td>39.3%</td>
</tr>
</tbody>
</table>

* = Significant at 0.05 level  
** = Significant at 0.01 level  
*** = Significant at 0.001 level

her rates for all these risk factors when compared to violent offenders. For the sub sample of youth with PCL-YV scores, while none of the findings are statistically significant, in all cases, except for the reporting of at least one family member abusing drugs or suffering from a mental disorder, the psychopathic offenders report higher rates for all these risk factors.

Another set of risk factors commonly associated with violence is the school environment and academic performance, specifically early and repetitive school failure. The youth in our sample are characterized by early failure in school and poor school performance. Specifically, 53.2% of the violent offenders and 45.5% of the non-violent offenders were not enrolled in school at the time that they committed their current offence. For both samples, on average, the last grade that was successfully completed was Grade 8. As the mean age of our sample is 16 years old, the youth in this sample are typically two academic grades behind the norm. It is also interesting to note the mean number of times these youth have changed schools when not required to because of graduations or normal grade advancement. Both violent and non-violent offenders have changed schools 6 times due to personal or family mobility, or problem or disruptive behavior, such as selling drugs at school, truancy, or getting into physical fights with students, teachers, or administrators. Moreover, 95% of violent and non-violent youth have routinized skipping school and began this behavior between 12 and 13 years old. Almost all of the youth have also gotten into trouble at school (95%), and the mean age at which they first get into trouble at school is 10 years old for violent offenders and 9 _ years old for non-violent offenders. Again, it is important to note that getting into trouble at school was operationalized as enga-
ving in behavior that could result in either a suspension or expulsion. The findings are quite similar when compared to the sub samples of psychopaths to non-psychopaths. In terms of school enrollment, 50% of psychopaths and non-psychopaths were enrolled in school. As with the larger sample, the mean last grade completed was 8 for both psychopaths and non-psychopaths. There are also extremely high rates of psychopaths (98%) and non-psychopaths (93%) getting into trouble at school. While not statistically significant, psychopaths began getting into serious trouble at school at 9.3 years old, while their non-psychopathic counterparts were 10.1 years old. Both psychopaths and non-psychopaths have changed schools 6.5 times.

Another important risk factor and predictor of violent offending and chronic offending is drug and alcohol use. Table 3 indicates the percentage of youth who self-report drug and alcohol use with a frequency of at least once a week. This frequency distinction is used because the rate of drug use and alcohol consumption without a consideration of frequency approaches 100% in this sample. Keeping in mind the frequency cutoff, this sample demonstrates a high rate of serious drug use and a large diversity of drug use.

Table 3. Drug use: frequency of at least once a week

<table>
<thead>
<tr>
<th></th>
<th>Non-Violent</th>
<th>Violent Offenders</th>
<th>Psychopaths (n=51)</th>
<th>Non-Psychopaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol</td>
<td>62.1%</td>
<td>69.5%</td>
<td>86.3%</td>
<td>83.3%</td>
</tr>
<tr>
<td>Marijuana</td>
<td>87.9%</td>
<td>84.2%</td>
<td>94.1%</td>
<td>91.1%</td>
</tr>
<tr>
<td>Heroin</td>
<td>52.6%</td>
<td>38%</td>
<td>23.5%</td>
<td>30.5%</td>
</tr>
<tr>
<td>Crack</td>
<td>71.7%</td>
<td>54.4%</td>
<td>37.3%</td>
<td>43.5%</td>
</tr>
<tr>
<td>Cocaine</td>
<td>50%</td>
<td>47.1%</td>
<td>47.1%</td>
<td>53.7%</td>
</tr>
</tbody>
</table>

* = Significant at .05 level
** = Significant at .01 level
*** = Significant at .001 level

Of great concern is the number of youth who report using hard drugs, such as crack and heroin, at least once a week. Again, it is interesting to note that a larger percentage of the non-violent youth in our sample report using hard drugs than the violent youth, although these differences are not statistically significant. There are also extremely high rates of marijuana use and alcohol use among the sub sample of psychopaths. It must also be noted that the onset of drug and alcohol use began at quite a young age for all the categories of offenders. Non-violent and violent offenders began using drugs at 11.9 years old and alcohol at 12 years old. Similarly, psychopathy offenders began using drugs at 11 years old and alcohol at 11.8 years old, while non-psychopathic offenders began using drugs and alcohol at 12 years old (p = 0.01).
Victims and offenders

There are also high rates of physical and sexual abuse suffered by the youth in this sample. In total, 19.7% of the violent offenders and 24.8% of the non-violent offenders report being victims of sexual abuse, and 43.5% of the violent offenders and 45.7% of the non-violent offenders report being the victim of physical abuse.

Table 4. Self-reported physical and sexual abuse

<table>
<thead>
<tr>
<th></th>
<th>Non-Violent Offenders (n=143)</th>
<th>Violent Offenders (n=364)</th>
<th>Psychopaths (n=51)</th>
<th>Non-Psychopaths (n=246)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Physical Abuse</td>
<td>31.7%***</td>
<td>15.5%*</td>
<td>29.1%***</td>
<td>13.9%***</td>
</tr>
<tr>
<td>Sexual Abuse</td>
<td>9.2%***</td>
<td>14.8%***</td>
<td>10.1%***</td>
<td>10.8%***</td>
</tr>
</tbody>
</table>

* = Significant at 0.05 level  
** = Significant at 0.01 level  
*** = Significant at 0.001 level

Males are more often the victims of physical abuse when compared to their female counterparts and non-violent offenders have a slightly, but statistically significant, higher rate of being the victim of physical abuse. While not statistically significant, nearly half of the males who met the threshold of psychopathy report being the victim of physical abuse, and nearly a fifth report being the victim of sexual abuse. For the female psychopaths, 10% report being the victim of physical abuse and 6.1% report being the victim of sexual abuse.

In terms of the presence of a mental illness or disorder, we are in the process of collecting official mental health information for the entire sample. However, we have included a self-report measure for Attention Deficit Hyperactivity Disorder (ADHD) in our research instrument. While not statistically significant, 30.5% of the non-violent offenders and 22.5% of the violent offenders self-report the presence of ADHD. In the PCL-YV sub sample, 36.2% of the psychopaths and 23.4% of the non-psychopaths self-report ADHD.

It appears that serious young offenders, whether violent or non-violent, have many of the individual, family, and educational problems identified with the life-course persistent adolescent pathways to criminality. And further, that psychopathic young offenders exhibit the more violent criminal histories, more time in prison, and have more family dysfunction than non-psychopaths. In effect, these data support theories that identify psychopathic traits as important in the explanation of a substantial proportion of violent young offenders.
REFERENCES


Victricks and offenders


The role of animal cruelty in the prediction of dangerousness

CHRIS MEDWAY AND DENNIS HOWITT

I | INTRODUCTION

Until recently, the possible link between animal cruelty and the various forms of domestic violence has been largely overlooked. In forensic psychology, it has been argued that animal cruelty is a common childhood characteristic of serial killers (Lockwood and Hodge, 1986). Animal cruelty has been suggested as a diagnostic sign of antisocial personality disorder (Felthous and Kellert, 1987). However, these have been rather isolated exceptions to the general lack of interest in the matter. There are signs of increasing research activity and there is at last some research data available especially since the publication of Lockwood and Ascione’s (1998) compendium book on cruelty to animals and interpersonal violence.

The belief of a link between animal cruelty and violence against people has been used as a call to action by some animal charities in the United States and elsewhere. For example, the Humane Society of the USA employs apparently authoritative statements such as “While not everybody who abuses animals will become a serial killer, virtually every serial killer first abused animals.” (Humane Society of the USA, 2002). The underlying message would seem to be that we should extend our concern about violence to animal cruelty since in this way humans may also be further protected. There are obvious inherent dangers in making a general case from extreme situations. While it may be the case that many extremely violent adults have a history of extreme or bizarre acts of cruelty to animals, such extreme cases reveal little about interpersonal violence and animal cruelty in general. Merely replicating studies of the association on further and varied populations needs to be supplemented by attempts to understand why there should be such a link. Despite the numerous studies of the childhood origins of crime and violent crime, we have very little information of animal cruelty as an
Victims and offenders

aspect of this. One exception is mentioned by Conboy-Hill (2000) who describes evidence that about a quarter of offenders imprisoned for violence admitted to cruelty against animals in childhood – this was five times the rate for non-violent prisoners.

2 TWO DIFFERENT VIEWS

Similarly, despite the voluminous research and theoretical literature on domestic violence, little is to be found about violence towards domestic pets in the context of family violence. Nevertheless, within the domestic violence field, one can find an alternative to the view that animal cruelty is symptomatic of individual pathology. Ganley (1989) described how violence against domestic pets is an aspect of, and integral to, partner or spousal violence and domestic violence in general. That is, she suggests that violence against domestic animals may be the result of the attempts of males to exercise power and dominance over their female partners. In this way, it is not meaningful or necessary to seek distinctive explanations of violence against domestic pets.

So two different views of animal cruelty are encompassed in the wider literature on violence. The first might be construed as being based on the view that pathological individuals are responsible for animal cruelty and extreme violence against people. The second perspective is based on the feminist conception of gender politics or gender socialisation. At this stage in our understanding, it should be stressed that these are complimentary views rather than competing ones. Both may be correct and feasible up to a point. This notwithstanding, assessing the relative strength and weakness of the two in explaining animal cruelty is important for the development of an adequate empirical and theoretical base. Furthermore, the context of the research may favour different points of view. Research on serial killers and animal cruelty may tell us little about domestic violence and animal cruelty, and vice versa. In the present paper we are concentrating on cruelty against domestic pets. There are good reasons for this - there is extensive violence within the family and contact with animals is primarily in the home for large numbers of individuals. We therefore need to be cautious about the relevance of findings to the general case.

3 RESEARCH

Possibly the most seminal research in the field is that of Ascione (1998). Central to this was a survey in the USA of women in domestic violence refuges. In this study, women who had suffered domestic violence were asked a number of questions. These concerned their own personal abuse by their partner and abuse to their domestic pets. Crucially, it was found that there were high levels of the co-occurrence of partner abuse and abuse of domestic pets within the household. If there was a pet in a household, there was a 71% likelihood that there had been
abuse of domestic pets. This figure includes threats to abuse the pets as well as actual violence against the pet. Fifty seven % of these cases involved actual violence against the animals, not just threats. In other words, the co-occurrence of the two forms of domestic violence seems to be high compared to some research on the co-occurrence of other forms of domestic violence. Why should there be such a trend?

In 2000-1 we carried out a study of the relationship between domestic violence, specifically partner abuse, and domestic pet abuse. This was in part a replication of Ascioni’s (1998) study. Again it involved women in domestic violence shelters. We incorporated into our study a matched sample of women from the community as a comparison. Participants completed with the aid of a third person Straus’s Conflict Tactics Scale. This is among the most commonly used measures of domestic violence and has been used effectively in major surveys of domestic violence. It measures domestic violence in an indirect way by asking participants how conflicts are resolved domestically without directly asking about abuse. Ways of dealing with conflict range from talking things over to the use of guns, for example. Our participants also completed a modified version of Ascioni questionnaire. Changes from the original included:

- The separation of hurting the pet from killing the pet;
- The inclusion of items to assess the use of pet abuse as a means of controlling the woman;
- Questions about the childhood history of their partner in terms of suffering and witnessing abuse of himself, his mother and pets.

The domestic refuge sample was 19 women and we created a matched sample of women living in the community in terms of age and employment status.

The two samples differed clearly in terms of the Straus Conflict Tactics Scale. The effect size or point-biserial correlation was 0.9 for two aspects of this measure that we had identified through factor analysis – the use of emotional tactics and the use of physical tactics. In other words, the two samples were very clearly differentiated in terms of spousal abuse by this measure. In terms of the co-occurrence of pet abuse and partner abuse, the data were very clear although the extent of this depended somewhat on the measure of pet abuse under consideration. The most graphic way of presenting the findings is to simply cross-tabulate spousal abuse by the reporting of any form of pet abuse – killing the pet, hurting the pet and threatening to harm the pet were our measures. Seventy-nine % of the domestic violence refuge sample reported at least one of these different forms. Only 21% reported none of these forms of abuse. In contrast, in the community sample, only 11% reported any of the three forms of pet abuse and 89% reported no form of pet abuse. If we exclude threat of pet abuse and concentrate on the measures of hurting or killing the pet, then the figures reduce somewhat. Fifty-eight percent of the domestic violence sample reported at least one of hurting or killing whereas only 5% of the community sample did. Killing of the pet
was actually quite rare at 11% of the refuge sample but none of the community sample. It is evident from this that there is a substantial degree of co-occurrence of partner abuse and domestic pet abuse. However, the difference between the two samples is greatest when threats of violence against pets are included in the analysis together with actual violence. Of course, threats of violence against domestic pets do not attract the legal considerations that threats of violence against one’s partner would.

All of the men in the domestic violence sample were, of course, abusive of their partners. We have also seen that some of them were also physically abusive of domestic pets. Were the men also violent outside of the home? This is an important issue in that if they were also violent outside the home then violence would seem to be more integral to their personalities. There was a tendency for the men who were violently abusive to pets to also act violently outside of the home. Of the men who were not violent outside of the home, 25% actually hurt a pet. Of the men who were violent outside of the home, 82% actually hurt a pet (Fisher exact test, p = 0.024). Although both of the men who killed pets in this sample were also violent outside of the home, this did not reach statistical significance. There was no trend for men who were violent outside the home to threaten to harm pets compared to men who were not violent outside of the home. Thus a substantial portion of men who use direct violence against pets were violent beyond the domestic situation.

The data on the men’s childhood were illuminating. In terms of being brought up in an abusive household, there was increased risk of threatening to abuse the pet. Fifty-three % of partners of women in the domestic violence sample had been brought up in abusive homes. However, 80% of the men who hurt the pet had been brought up in abusive homes compared with only 20% of men who had not been brought up in abusive homes (Fisher Exact Test, marginally significant p = 0.07). Both the men who had killed pets were reported as being brought up in abusive homes. The trends were very much the same for the question that asked about the abusers witnessing violence as a child. Seventy five percent of those who had witnessed violent abuse as a child actually hurt a pet whereas only 29% of those who did not witness abuse actually hurt a pet (Fisher Exact Test, marginally significant, p = 0.07). One suggestion is that the abusive control of women is being learnt rather than the abuse of pets as such. Personal violent victimisation as a child, however, did not appear to be associated with greater levels of pet abuse. Thirty seven % of the domestic violence sample was reported to have been subjected to violence in childhood. However, this did not relate significantly to whether they used threats of violence to the pets and actual violence to the pets. Both men who killed pets had been subjected to violence in childhood. The variables concerning childhood abuse of pets or witnessing the abuse of pets demonstrated little or no co-occurrence with adult abuse of pet variables. This was because virtually none of the men were reported to have abused or witness the abuse of pets in childhood. Our research strategy, it should be stressed, relies on the information provided by the battered spouses and so should be regarded with some caution on such matters.
The women’s experiences provided further evidence that dominance and control issues are an important component of why domestic animals are treated cruelly. Nearly all of the domestic violence sample (95%) thought that violence to them preceded violence to the pet animal. Furthermore, 82% of women whose pets had been hurt said that it was in an attempt to control them. There were a number of informative associations between pet abuse and the feelings of the women. One of our additional questions was ‘How close were you to the abused pet?’ There were significant associations (Fisher Exact Test, p. = 0.01) between being close to the pet and threatening to harm the pet and actually harming the pet. There was some evidence that concerns about the safety of the pet discouraged the women from leaving the abusive environment. The questionnaire included an item ‘Did concern for your pet delay you leaving home?’. Women whose partner was threatening to harm the pet tended to delay seeking the safety of a refuge because of concerns about the pet (Fisher Exact Probability, p. = 0.04). It is worth noting that this tendency was not found for the women whose partner’s actually harmed the pet (Fisher Exact Probability p. = 0.38). This may imply that the threat to the pet was primarily a means of controlling the woman – discouraging her from leaving.

4 | CONCLUSION

What appears to emerge out of these data is evidence of a need to consider both explanations of animal cruelty - that it is the result of a violence prone personality and that it is aspects of an attempt to control women - have a degree of validity and support. Neither can be totally accepted or rejected. It would appear that violence towards and the threat of violence towards domestic pets is a characteristic feature of many domestic violence situations. Nevertheless, the situations are first and foremost ones of violence against the female partner. This almost always precedes violence against the domestic pets. Women experience violence against domestic pets as part of an attempt to control the woman. The threat of violence and violence against domestic pets seem to be different since they occur in somewhat distinct groups of men. Men with a tendency to violence outside of the home are the ones most likely to use violence - threat of violence against pets is more characteristic of men whose violence is confined to their partner. The aetiology of violence against pet animals in childhood seems to have unexpected features. Abusive households are fairly typical of the childhood of these men. However, witnessing violence against ones mother seems to be more influential than violence perpetrated against themselves in childhood. Similarly, witnessing violence against domestic pets in childhood is unrelated to abusing pets as adults largely because evidence of such experiences is rare.

The practical significance of these findings is difficult to assess. A simple practical step of permitting domestic pets in refuges (or providing adequate provision of the pets in some other way) might facilitate the women leaving home sooner and, equally, reduce the period of risk to and abuse of the pet. Increased co-ope-
rations between agencies for humans and agencies for animals might ensure that provision is made to take care of the safety of animals as part of the case review. It would seem to be less clear the extent to which known cruelty to animals might be helpful in preventing abuse to partners. Research into the animal abuse cases known to animal protection agencies would cast light into its potential usefulness for interventions into family violence.

REFERENCES


Development of criminal careers in psychopathic and non-psychopathic young offenders

SPIDEL, A. GREAVES, C., HERVÉ, H. F & HARE, R. D.

I | INTRODUCTION

Psychopathy is a socially devastating personality disorder comprised of a unique constellation of affective, interpersonal, and behavioural characteristics (Cleckley, 1976; Hare, 1991). Affectively, psychopaths are callous with regard to the feelings and rights of others, manipulative, egocentric, and lack a normal sense of empathy, guilt, and remorse (Hare, 1998). Behaviorally, psychopaths are impulsive, have poor behavioural controls, and disregard social norms and expectations. Despite the fact that they are known to be more deviant and problematic than non-psychopaths, as a result of negative ethical and practical considerations (e.g., adverse effects of negative labeling, controvertible diagnoses; see Edens, Skeem, Cruise, & Cauffman, 2001; Quay, 1987) the DSM-IV (American Psychological Association, 1994) is cautious in assigning diagnoses to children, especially regarding psychopathic personality disorder. Such a diagnosis will only be given if the related traits are trans-situational, enduring, and unrelated to the child’s particular developmental stage (APA, 1994). The psychopathic personality does not simply appear in adulthood however; the developmental paths to antisocial behaviours and personality traits exist very early in life (Farrington, 1997).

While research on adult psychopathy continues to expand, there has been considerably less research devoted to the construct of psychopathy as exhibited in youth, particularly with female adolescents. As such, the present study investigated those variables largely overlooked in the literature, such as specific demographic, criminal, and deviant characteristics as manifested in both male and female adolescent psychopaths.
Given the stability of personality traits over time, it is a reasonable prediction that at least some correlates of adult psychopathy will be found within psychopathic adolescents. Although not all psychopaths commit violent acts, the construct of psychopathy is closely related to violence (Hare, 1991). This is likely due to psychopaths’ lack of affect, guilt, or remorse – all of which enable nonpsychopaths to prevent/suppress violent and aggressive behavior. As a result, while psychopaths represent only 15-25% of incarcerated adult offenders, they are responsible for a disproportionate amount of crime (Hemphill, Hare, & Wong, 1998). Adult psychopaths are more likely than non-psychopaths to be involved in crime (Wong, 1984; Hare & Jutai, 1983; Kosson, Smith, & Newman, 1990) and to commit a wider variety of offence types (Kosson et al., 1990). Considering psychopathy has emerged as a robust predictor of deviant and violent behaviour in adult offenders, extending the concept to youth has clear utility and has been demonstrated in a small number of studies.

Hare (1991) and Wong (1984), among others, have found that among psychopaths, criminal activities begin relatively early in life. Compared to non-psychopathic offenders, formal contacts with the criminal justice system begin several years earlier, around the age of 15, for psychopathic offenders. Even more important is that the rate of psychopathy in incarcerated adolescent males is estimated to be 34% (Forth, Kosson, & Hare, 1996; Hemphill, Hart, & Hare, 1994) – a rate greater than that found in adult offenders. Moreover, Gretton’s (1998) study of adolescent offenders demonstrated that the proclivity towards violent acts continued into adulthood. In terms of studies addressing females and violence, Strachan (1991) found that one-third of his sample of female offenders were assessed as psychopaths. Furthermore, these women were responsible for approximately half of both the violent and non-violent offences committed by the entire sample (Strachan, 1991). There is considerably less literature on psychopathic females and their connection to violence, and the findings that do exist are inconsistent (Stattin & Magnusson, 1989).

When determining propensity for future violent behaviour, frequency, recency, and severity of past like behaviours should be considered. As well, the youth’s social milieu plays a major role in criminality. Aggressive youth form antisocial peer connections (Cairns, Cairns, Neckerman, Gest, & Gariepy, 1988). Thus, associations with other delinquent youth become a risk factor for delinquency and violence, continued use of violence, and both overt and covert delinquent activities (Keenan, Loeber, Zhang, Stouthamer-Loeb, & Van Kammen, 1995). Gang membership is related to serious and violent criminal activities. Upon entry, rates of youth violence increase, remaining so until the individual breaks his/her affiliation (Thornberry, Krohn, Lizotte, & Chard-Wierschem, 1993). Although gang membership has been investigated in general in adolescents, no specific literature has addressed the link to psychopathy or the link with female adolescents.

Similarly, as with gang membership, substance abuse consistently appears as a risk factor in adolescent violence (Loeber & Dishion, 1983; Loeber & Hay, 1997;
Loeber & Stouthamer-Loeber, 1987). Substance use has also been implicated in recidivism (Dembo, Turner, Chin Sue, Schmeidler, Borden, & Manning, 1995). For example, Brook, Whiteman, Finch, & Cohen (1996) conducted a 20-year longitudinal survey and found that adolescent drug use was related to concurrent and future delinquency. In addition, Smith and Newmann (1990) found that compared to non-psychopaths, psychopaths were more likely to have a substance-related disorder, abuse specific types of drugs, and use multiple drugs. There was also a significant association between psychopathy and offenders’ age of first intoxication, as well as age at first arrest. Hart & Hare (1989) found a significant association between drug dependence/abuse diagnoses and psychopathy.

Through an examination of male psychopathic adolescents and substance use, Mailloux, Forth, & Kroner (1997) found psychopathy was significantly correlated with an earlier age for alcohol and drug use onset, and a greater number of drugs tried. Brandt, Kennedy, Patrick, & Curtain (1997), however, found no association between adolescent psychopathy and substance use, contradicting previous findings, as well as those in the adult literature. These contradictions may be explained by the comparatively small sample size, and the fact that they used a less sensitive measure of substance abuse (Brandt et al., 1997). Reviewing this literature, Hemphill et al. (1994) conclude that there is at least a moderate correlation between drug dependence/abuse and psychopathy in male incarcerated offenders (also see Forth & Mailloux, 2000). Interestingly, an equal relationship was found between males and females for physical assaults while under the influence of illicit substances (Dukarm, Byrd, Auinger, & Weitzman, 1996). The association between females and substance abuse has not been extended further, particularly with reference to psychopathy.

In light of the preliminary and contradictory findings with adolescents and specific criminal characteristics, the current study was designed to add to the literature by investigating the link between psychopathy, crime, gang affiliation, and substance abuse in male and female adolescents. This study employed a self-report measure, the Criminal Profile Questionnaire (Hervé & Spidel, 1999) to assess characteristics of psychopathic juvenile offenders. One such characteristic was the relationship between psychopathy and certain criminal behaviours. It was expected that psychopathic individuals would commit more crimes and therefore spend more time in jail and have more convictions than non-psychopaths. In addition the link between psychopathy and gang membership was addressed. As psychopaths tend to dominate group environments and impose their beliefs on others (Hare, 1993), it was hypothesized that psychopaths would tend to hold a high rank in gangs. The association between psychopathy and a propensity for violent acts, as well as gang activity and sustained violent activity, lead to the prediction that psychopaths in this sample would have spent more time involved in gang activity. The last area of focus was on psychopathy and substance abuse. It was hypothesized that in concordance with the literature, psychopaths as compared to non-psychopaths would have tried, and would perhaps currently be using, a greater variety of illicit substances.
2 | METHODS

2.1 | PARTICIPANTS

The study was conducted at the inpatient assessment unit of the Youth Forensic Psychiatric Services, a co-ed youth facility in Burnaby, British Columbia, between 1999 and 2001. To be held in this facility, the juvenile must have been referred for a psychological assessment as a result of criminal activity and be between the ages of 12 and 19. At the time of the study, participants were either on remand for a pre-trial assessment regarding recent charges, or were serving a custodial sentence.

The sample consisted of 33 Canadian young offenders, 21 male and 12 female, with an average age of 15.5 (SD = 2.33), at the time of recruitment. The participants were approximately 61% Caucasian, 12% First Nations, 12% Asian, and 3% African.

To be included in the sample, the offender had to have undergone a psychological/psychiatric assessment at the Youth Forensic Psychological Services (YFPS) as the files written during these assessments were the principal source of information for this study. In addition, offenders had to be at least 12 years old, with sufficient reading ability and no severe psychotic symptoms as determined by the staff on the unit.

To maintain participant confidentiality, anonymity and identity protection under the Young Offenders’ Act, participants were assigned a code number and the list of names were kept securely in the laboratory, separate from the data. At the end of the study, all identifying information was destroyed.

2.2 | PARTICIPANTS

The Hare Psychopathy Checklist-Revised (PCL-R; Hare, 1991) is currently regarded as the ‘gold standard’ for operationalizing the construct of psychopathy in adults. The 20-item checklist assesses the affective/interpersonal and behavioural correlates of the disorder. The score from 0-40 reflects the extent to which the individual represents the prototypical psychopath. Two stable factors have emerged through factor analysis of the PCL-R (Hare et al., 1990; Harpur, Hakstian, & Hare, 1988). Items loading on Factor 1 reflect affective/interpersonal traits (e.g. shallow emotions, lack of remorse and empathy, selfishness, callousness) considered being core psychopathic personality traits. Items loading on Factor 2 reflect antisocial behavioural characteristics (e.g. impulsivity, criminal versatility, poor behavioural controls). There is an accumulation of literature supporting both the reliability and validity of the PCL-R (e.g. Hare, 1991; Hare, Forth, & Strachan, 1990; Harpur, Hare, & Hakstian, 1989) in a variety of samples (e.g. Hart, Hare, & Harpur, 1992; Hobson & Shine, 1998; Cooke & Mitche, 1999; Molto, Poy,
Torrubia, 2000). Its ability to predict violent recidivism has also been substantiated (e.g. Rice, Harris, & Quinsey, 1990; Harris, Rice, & Cormier, 1991; Hemphill, 1991; Serin, 1991).

While the PCL-R items are appropriate for adult assessments, changes were necessary to account for an adolescent’s relatively limited life experience in order to extend age-appropriate assessments. The PCL: Youth Version (PCL:YV; Forth, Kosson, & Hare, 1995) is a modification of the well-established Hare Psychopathy Checklist-Revised (PCL-R; Hare, 1991). The PCL:YV is comprised of the PCL-R’s 20 items, three of which are adapted for the assessment of adolescents, item 17 (Many short-term marital relationships), items 18 (Juvenile delinquency) and item 20 (Criminal versatility). The PCL:YV items are scored and totaled in the same manner as the PCL-R. The PCL:YV has demonstrated high internal consistency and interrater reliability (Brandt et al., 1997; Forth, 1995; Gretton, 1998; McBride, 1998). Factor analysis of the PCL:YV showed two correlated factors parallel to those in the literature with adults with the PCL-R (Brandt et al., 1997; Forth, 1995; Forth et al., 1996). The value of the comparative fit index was .83 when compared with Harpur et al. (1988). In addition the congruence coefficient ranged from .91 to .89 for factor 1 and .84 to .89 for factor 2 (Brandt et al., 1997).

Use of the PCL:YV to assess psychopathy is supported by a growing body of research showing its reliability and validity in adolescent offender populations (Brandt, Kennedy, Patrick, & Curtin, 1997; Forth, Hart, & Hare, 1990; Gretton, 1998; Mailloux, Forth, & Kroner, 1997; McBride, 1998). It has shown high interrater reliability (alpha = .87) and good internal consistency (alpha + .85) (Brandt et al., 1997; Forth et al., 1990; McBride, 1998).

In the present study, a senior undergraduate student conducted PCL:YV assessments on all participants. A graduate student independently rated 8 of the participants to evaluate interrater reliability. The two raters had received training mainly with the Hare PCL-R. The training consists of workshops and scoring of training cases until the scores fall within three points of the senior trainers’ scores, for 10 cases.

Assessments were based on file information. The file-only method is also acceptable according to the PCL:YV rating guide and reliable and valid judgments can be made from file information only as long as the information is varied (see Grann, Langstrom, Tengstrom, & Stalenheim, 1998; Rice & Harris, 1997; Wong, 1984, 1988) as was the case here. For the purposes of this study, PCL:YV assessments were used categorically. Although the PCL:YV is a dimensional measure of psychopathy, Forth & Mailloux (2000) state that a categorical diagnosis is acceptable for research purposes. Accordingly, our sample was divided into High (H: n = 11), and Low (L: n = 22) PCL:YV groups, using the adult version recommended cut-off of 30 for the High group and 20 for the Low group (see Hare, 1991).
2.3 | CRIMINAL HISTORY

The Criminal Profile Questionnaire (Hervé & Spidel, 1999) is a self-report questionnaire consisting of age-related questions regarding various antisocial behaviors. The first section relates to drug use behaviors including drugs tried, and Likert scale type ratings of the typical amount of usage, and of how much they believe the substance contributed to their criminal activities. Their involvement in, start date, and frequency of non-criminal behaviours (e.g. lying, bullying, sexual activities, school disruptions) are also recorded. Criminal background information self-reported includes date and type of first crime, motivations, influences of substances, number of crimes committed for which they were caught and not caught, and time spent in custody. Gang affiliation questions include date joined, their role as leader or follower, number and age range of members, if they have ever changed gangs, and the proportion of crimes committed when with other gang members compared to alone. Three graphs to plot their personal criminal history were provided. The first graph is to plot how many crimes they committed per year, the second for the number of violent crimes per year, and the third for the number of non-violent crimes per year. Finally, they were to complete summary sheets to describe their last four crimes (caught and not caught) in reverse chronological order. Descriptions included information on crime type, date, length, witnesses, accomplices, consequences, planning, victims and weapons. This, as with most self-report questionnaires, is susceptible to reporting biases. However, as children become adolescents, their self-report is more valid and reliable with regards to antisocial behaviours. Further, such behaviours may not be witnessed by parents or other related adults (Kamphaus & Frick, 1996). There are, however, no validity scales for this questionnaire.

2.4 | PROCEDURE

Participants volunteered to be a part of the study, and signed a consent form allowing researchers to review their institutional files. Subjects were given the Criminal Profile Questionnaire to complete at their own pace, and the researchers clarified questions as needed. Each subject was paid with two candy bars upon completion of the CPQ. Data obtained for the PCL:YV scoring was based on file history information. This included nursing observations, recorded several times a day, police information such as witnesses’ statements, arrest reports, investigation reports, etc, reports written by a social worker who garnered information from family members, school teachers and therapists, reports of the psychiatrist and/or psychologist, victim statements, progress summaries, and demographic information.
3 | RESULTS

3.1 | PSYCHOPATHY

A subsample of 8 (24.2%) participants was assessed on the PCL-YV by two raters. Intraclass correlation coefficients were .97, p < .000 for a single rating and .98, p < .000 for the average of two ratings. Scores on the PCL-YV ranged from 9 to 34 with a mean of 24.2 (SD = 6.9). These scores are comparable with those obtained in other similar samples (e.g., Brandt et al., 1997; Forth et al., 1990).

Groups were compared via two-tailed t-tests, with homogeneity of variance assessed via Levene’s test and, if significant, corrected via Welch’s test. As compared to Group L, Group H reported committing a greater number of crimes for which they were not caught (t(1, 28)=6.139, p<.02) but no differences emerged in relation to criminal convictions (t(1,29)=2.168, p>.15). Nonetheless, Group H spent a significantly smaller amount of time in jail than Group L (t(1, 19)=5.855, p<.025). In addition, they were more likely to belong to a gang (t(1, 31)=5.855, p<.025), and spend a longer period of time involved in gangs (t(1, 8)=6.742, p<.05). However, there was no difference in regards to the position the offender played in the gang (t(1,9)=.318, p>.55). Group H used significantly more categories of drugs than did Group L, (t(1, 29)=6.015, p< .025). However, the average drug use level was not significantly different than that of Group L (t(1, 26)=.209, p>.65) and the impact of drug use on offending was reportedly the same for both groups (t(1, 25)=.051, p>.9).

Table 1. Means and Standard Deviations (SD) for Psychopaths versus Non-psychopaths offenders across criminal history variables

<table>
<thead>
<tr>
<th></th>
<th>Psychopath Mean</th>
<th>Psychopath SD</th>
<th>Non-psychopath Mean</th>
<th>Non-psychopath SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories of drugs</td>
<td>1.57</td>
<td>.98</td>
<td>3.18</td>
<td>1.97</td>
</tr>
<tr>
<td>Drug use level</td>
<td>2.90</td>
<td>1.39</td>
<td>3.18</td>
<td>1.20</td>
</tr>
<tr>
<td>Drug use impact on offending</td>
<td>2.20</td>
<td>1.79</td>
<td>2.35</td>
<td>1.22</td>
</tr>
<tr>
<td>Number of crimes not caught</td>
<td>55.00</td>
<td>26.93</td>
<td>22.32</td>
<td>31.32</td>
</tr>
<tr>
<td>Number of criminal convictions</td>
<td>8.43</td>
<td>3.95</td>
<td>4.83</td>
<td>6.06</td>
</tr>
<tr>
<td>Gang affiliation</td>
<td>0.86</td>
<td>0.38</td>
<td>0.29</td>
<td>0.46</td>
</tr>
<tr>
<td>Time involved in gang</td>
<td>24.00</td>
<td>0.00</td>
<td>14.00</td>
<td>14.14</td>
</tr>
<tr>
<td>Gang position</td>
<td>2.00</td>
<td>1.00</td>
<td>2.43</td>
<td>1.13</td>
</tr>
<tr>
<td>Time in prison</td>
<td>1.17</td>
<td>2.86</td>
<td>42.00</td>
<td>28.85</td>
</tr>
</tbody>
</table>
DISCUSSION

The goal of the present study was to identify differences in the criminal activities of psychopathic verses non-psychopathic juvenile offenders using the self-report Criminal Profile Questionnaire (Hervé & Spidel, 1999) while validating a construct of psychopathy in male and female adolescents. As expected, some notable and significant differences were found between high and low PCL:YV scorers.

This study determined that non-psychopathic youth spent more time in prison than their psychopathic counterparts. Perhaps like their adult equivalents, juvenile psychopaths are able to con and manipulate criminal justice professionals in believing that they are responding well to treatment and, therefore, ready for early release (Hare, Forth, & Hart, 1989). Psychopaths are considered ‘natural liars’ as they have an aptitude and propensity towards deception (Ekman, 1985). In a study examining deceptive motivations in adolescent offenders, Spidel, Hervé, Greaves & Hare (2002) found that, related to the present findings, psychopaths lied most often to avoid punishment, to obtain a reward, and to heighten self-presentation. Psychopaths do not have a guilty conscience that would normally inhibit lying, and can be glib and convincing, especially when they stand to gain something (i.e. their freedom). Moreover, they have an ability to judge and manipulate people and situations, and use this information to their advantage (Hare, 1991, 1993). Their pervasive tendency to lie no matter the situation will present itself especially in circumstances regarding manipulation of the justice system with which they are often involved. Another facet of psychopaths’ manipulative tendency may be demonstrated by the fact that psychopaths committed more crimes for which they were not caught. This finding suggests that psychopathic juveniles are perhaps better able to escape detection during and/or after the commission of their crime(s). An alternative explanation is that they are capable of convincing others (e.g., fellow gang members) to take criminal responsibility. This hypothesis is further substantiated by the findings here relating to gang affiliation.

Psychopaths in this sample had a higher gang membership rate, and stayed involved in gangs for a longer period of time than did the non-psychopaths. Although contrary to our hypothesis that psychopaths would be more likely to be a leader in a gang, psychopathy was not associated with the position held within gangs. However, it is unclear as to whether this is the actual case or if it is due to psychopath’s tendency to lie or attempt to project a good impression. As such, an explicit association between psychopathy and gang involvement was unclear. However, these findings might indicate that psychopaths are likely to thrive in gang environments through manipulation of other members, and the dispersion of responsibility for criminal acts. Psychopaths have a tendency to externalize responsibility for their criminal acts even in situations where there is no specific other to blame (Harry, 1992). Conning and deceit are used to execute personal objectives, as is remorseless manipulation of others (Hare, 1991). Both psychopathy and gang membership are associated with a higher rate of serious and
more violent criminal behaviour (Hart & Hare, 1997; Thornberry, Krohn, Lizotte, & Chard-Wierschem, 1993). Given this, and the psychopath’s propensity for manipulation and use of others, a gang association could be conducive to the lifestyle of some criminal psychopaths, and may especially work to the advantage of young offenders. Those who become involved will most likely find that group offending dynamics facilitates their antisocial behaviour. As such, this may be an ideal environment within which to hone their manipulation and deception skills while having a decreased risk of having to take sole responsibility of their violent offences.

In terms of substance use factors, average drug use level was not significantly different across groups, nor was the impact of drug use on offending. Consistent with past findings (Mailloux et al., 1997), psychopathic adolescents were found to have used a wider variety of drugs than their non-psychopathic counterparts. This indiscriminant usage has been suggested to be associated with risk taking behaviour and impulsivity both implicated in psychopathy (Mailloux et al., 1997; Rutherford Alterman, Cacciola, 2000). Both psychopathy and substance use have been implicated in an increased risk for violent offending in youth (Forth et al., 2000; Loeber & Dishion, 1983; Loeber & Hay, 1997; Loeber & Stouthamer-Loeber, 1987) and as such has important implications for the justice system.

In terms of implications, the fact that psychopaths were caught less often for their crimes has risk assessment and other clinical implications. If we can expect that juvenile psychopaths are being charged less than non-psychopaths but are committing more crime, it follows that reliance on official records would be unadvisable. Additionally, we must consider this when assessing risk and the likelihood of future recidivism. Future violence is best predicted by prior violent behaviour (Farrington, 1991; Kohlberg, LaCrosse, & Ricks, 1972; Mossman, 1994; Parker & Asher, 1987; Tolan, Guerra, & Kendall, 1995), and the risk increases with a higher incidence of past violent behaviour. This may be even more pertinent for the psychopath. In addition, recognition of the distinct issues gang involvement brings to rehabilitation options and success is of importance. In terms of treatment it is important to address social factors and consider the strong influence of peer relations on the individual. Strategies must be implemented to aid the offender in dealing with peer pressure when released back to their environment. This may be of greater importance in adolescents as peer pressure is a factor that tends to impact adolescents more strongly than adults. Given that adult psychopath are resistant to current treatment options (Ogloff, Wong, & Greenwood, 1990; Rice, Harris, & Cormier, 1992), developing a focused treatment program may be the first course of action regarding an adolescent psychopath. In terms of substance abuse and treatment, both groups, psychopathic or not, will differ in their compliance and response to substance abuse treatment programs. In a study of adults in substance abuse treatment by Alterman, Rutherford, Cacciola, McKay, & Boardman (1997), psychopathy was a robust predictor of treatment non-completion. In addition, methadone patients were able to make improvements, but success was much harder for cocaine and benzodiaza-
Victims and offenders

Pine addicts. This is consistent with research findings on the etiology of drug and alcohol addiction that has focused on certain risk factors for the disorder (Conrod, Peterson, & Pihl, 1997). Moreover, in substance abusing women, it has been demonstrated that four personality risk factors (sensation-seeking, impulsivity, anxiety sensitivity and introversion/hopelessness) are differentially associated with drugs’ reinforcing affects (Conrod & Lodato, 2001; Conrod, Pihl, Stewart & Doniger, 2000) suggesting four separate etiologic mechanisms for substance related behaviour. Psychopaths have been described as specifically possessing at least two of these risk factors: sensation seeking and impulsive behaviour (Cleckley, 1976; Hare, 1991). Accordingly, it may be investigation into these areas that will uncover the link between specific substance and the psychopathic personality. As a result, examination as to whether these findings hold for adolescents is an essential area for further exploration.

Limitations of the current study include its relatively small sample size and the fact that our measure of criminal behaviour was based on self-report. These points should be addressed in future studies. The study was not however, constrained by the potential plasticity of psychopathy, as we focused solely on adolescent psychopathy. The direction of future studies should be towards tracking the developmental course of juvenile psychopathy to its outcomes in adulthood. In addition, a comparison of gender differences would be of interest.

In summary, identification of particular criminal activities of psychopathic youth adds to our general conceptualization of the disorder. This also highlights the potential for early intervention with young psychopathic offenders, and as a consequence, prevention of a considerable amount of victimization given their criminal persistence. As more information is gathered, individualized treatment plans tailored to the specific needs of the psychopathic juvenile have the potential to be developed. Despite disparate beliefs on the development and course of the disorder, given its serious nature, disregarding the construct of psychopathy in adolescence is not a feasible option considering the personal and financial costs for society.

REFERENCES


DEVELOPMENT OF CRIMINAL CAREERS IN PSYCHOPATHIC AND
NON-PSYCHOPATHIC YOUNG OFFENDERS


Victims and offenders


DEVELOPMENT OF CRIMINAL CAREERS IN PSYCHOPATHIC AND NON-PSYCHOPATHIC YOUNG OFFENDERS


Victims and offenders


DEVELOPMENT OF CRIMINAL CAREERS IN PSYCHOPATHIC AND NON-PSYCHOPATHIC YOUNG OFFENDERS


Psychopathy and deceptive motivations in young offenders

Alicia Spidel, Hugues F. Hervé, Caroline Greaves, Barry Cooper & Robert D. Hare

I | INTRODUCTION

Deception is a frequent occurrence in everyday life (DePaulo & Kirkendol, 1989), and its empirical study is of critical importance in the forensic context. The stakes for the individual and society are heightened when this group is involved, as their lies tend to have more significant consequences (e.g., influencing parole decisions). Moreover, some type of deception is typically an “essential prerequisite of most crimes” (Rogers & Cruise, 2000, p. 275). The present research was concerned with the motivations to deceive among juvenile offenders. The central focus was on identifying different deceptive motivations and examining their relationship to psychopathy.

To date, several studies have explored the motivations for deception in non-forensic samples (Camden, Motley & Wilson, 1984; Hample, 1980; Lippard, 1988; Turner, Edgley & Olmstead, 1975). Most of these studies have focused on white lies in an attempt to investigate social motivations. This research is not without flaws, particularly when researchers have attempted to generalize the findings to forensic populations. The participants used in these studies were undergraduates dissimulating about everyday situations (Camden et al., 1984; Hample, 1980; Lippard, 1988; Turner et al., 1975). As a result, the consequences of lying were typically minor and the liars were relatively unsophisticated (Lippard, 1988) compared with a forensic population. A further problem with the aforementioned studies is that the lies were repeatedly categorized as prosocial lies hence the motivations were generally social. Prosocial lies, or white lies, are false statements that help keep social interactions smooth and positive (Ford, 1996). Consequently, their impact on assessment of credibility, while important, is comparatively trivial with respect to lies revealed in a forensic sample (Petitclerc &
Victims and offenders

Hervé, 1999). Lies can generally be categorized in motivational terms as aggressive, defensive, or as white lies (Ford et al. 1988) and by pathology as normal or abnormal (Ford et al., 1988). The current investigation tends to focus on aggressive lies rather than white lies due to the population of focus. For example, a psychopath may be able to put on a good show for the parole board (Hare, Forth, & Hart, 1989) and thus achieve parole earlier than others (Hare et al., 1989). Another problem with the previous studies is that they were descriptive and relied on self-report. When dealing with offenders, particularly psychopaths, it is imperative not to solely rely on self-report; as they are unlikely to admit to the lies investigators are most interested in – the lies with substantial consequences for the offender if detected, and for society if undetected. As such, it has been suggested by Hare et al. (1989) that caution should be taken when using self-report in prison populations to assess for deception.

Three studies have investigated motives for deception in a forensic population. Recently, Rogers and Cruise (2000) conducted a study in order to evaluate differences in the deception-related motivations of offenders as a function of psychopathy. Rogers and Cruise analyzed an extensive data set based on previous studies of psychopathy (see Rogers & Cruise, 2000) and utilized a modified version of the Psychopathy Checklist Screening Version (PCL: SV; Hare, et al., 1989). They asked clinicians to rate the 58 subcriteria (17 were determined to address deception) in addition to the 12 criteria to assess deception. Of the 58 subcriteria, a principal axis factoring on the 17 subcriteria was performed and a three-factor model found which included: implausible presentation, characterized by an “unbelievable display with respect to statements and emotional expressions” (p.277), lies to con and manipulate the target, and lies to deny criminality, which refute criminal involvement and blame external sources. These motives are similar to ones found in models using non-forensic samples, with the exception of denying criminality. Implausible presentation is conceptually similar to lies to maintain self esteem (see Ford, 1996) and lies to win the admiration of others (see Ekman 1997). Conning and manipulate the target has parallels with lies to manipulate the behaviour of others (see Ford, 1996) and to obtain a reward (see Ekman, 1997). However, although this study employs a forensic sample, it does not address the motives for individual lie, rather its focus is on overall presentation of the offenders with specific application to psychopaths.

Another study that addressed motives for deception in psychopaths was a pilot study performed by Petitclerc, Hervé, Spidel, & Hare (2000). In this study the 3 categories established in Rogers and Cruise’s (2000) investigation (implausible presentation, lies to con and manipulate the target, and lies to deny criminality) were conceptually similar to motivations to heighten self-presentation, obtain a reward, and avoid punishment respectively [see Petitclerc and Hervé (1999) model below]. However, in addition to these findings duping delight was also found to be more characteristics of psychopath’s deception. In addition, this study was expanded (Spidel, 2002) to include more offenders and similar findings were revealed. This research with forensic samples, utilized Petitclerc and
Hervé (1999) model describing motivations for deception in offenders. Petitclerc and Hervé’s model has improved on previous frameworks in several ways. Although the lies are not specific to a forensic population they were delineated with a forensic population in mind. Given that lies provided in forensic contexts are of more diverse subject matter than typically associated with a sample of undergraduates, it is likely that they will have more variable motivations. Another advantage to this version is that it is based upon both clinical and research experience, as well as a review of the relevant literature, whereas previous types of motivation were specified from self-report, and were therefore descriptive. Petitclerc and Hervé’s model includes ten motivational categories. Briefly, they are as follows:

(a) Compulsive:
These lies seem to be completely lacking a purpose. They are usually not self-serving and, in fact, may be self-destructive as the deception is random and likely to be discovered. Compulsive lies are usually quite spontaneous (Ford, 1996). Those considered ‘pathological liars’ are known for their compulsiveness (Ekman, 1997); that is, they cannot control telling them. Several authors have suggested that psychopaths may fall into this category (Ekman, & Frank, 1993). Their lies, however, may be motivated by self-serving purposes, and therefore do not necessarily qualify for our “compulsive” category.

(b) Secretive:
A secretive lie is motivated by the offender’s desire to keep some personal information concealed (Ford, 1996). The offender is reluctant to give the target personal information, regardless of the latter’s desire or need to know the truth. Ford (1996) talks about lying “to preserve a sense of autonomy” (p. 88). As adults, “people who react strongly to control or intrusiveness from others may resort to lying in an effort to maintain a sense of independence” (p. 88).

(c) Avoid Punishment:
Lies in this motivational category are by definition self-serving lies. They can take the form of a general deceptive statement to evade a punitive consequence (Lippard, 1988), or as a lie to avoid a conflict, which is conceptualized by Petitclerc and Hervé (1999) to be a form of punishment. For obvious reasons, these are also probably the most frequently encountered types of lies in the criminal justice system. Ford (1996), Kropp, and Rogers (1993) propose that different types of individuals lie to avoid punishment due to various motivational pressures. Some individuals place their own needs and desires above the consequences of their lies, while others weigh the pros and cons of lying and telling the truth and reason that lying is the best way to cope with their present situation (Kropp & Rogers, 1993).

(d) Avoid Negative Evaluation:
This category includes lies concerning a topic that the offender is shameful or worried over being judged about. Such lies are said to occur when the offender is
Victims and offenders

mindful of, and is concerned about, the target’s opinion of him, or when he is
generally careful about self-presentation (Petitclerc & Hervé 1999). People with
low self-esteem may feel as though they are a failure when their talents and abili-
ties fall short of their expectations (Ford, 1996). In an attempt to regulate the
person may lie to close the gap (Ford, 1996) between their beliefs and reality.

(e) Protective:
Lies in this category are used in order to avoid the physical retaliation of another
person. Classification of a deceptive statement in this category means that its
purpose was not to avoid a legal or emotional punishment, but instead served to
avoid serious physical injury or death to the liar (Petitclerc & Hervé 1999).

(f) Obtain a Reward:
Lying to obtain a reward can be regarded as going beyond lying to avoid punish-
ment. Any gain from this lie is something undeserved, and would not have been
obtained by the offender under other circumstances (Petitclerc & Hervé 1999).
In this case, gains could be physical (e.g., obtaining sexual favors from a spouse),
situational (e.g., early release from punishment), material (e.g., money), or
internal, (e.g., attention). Ekman (1997) states that this is the second most often
mentioned reason for lying, after lying to avoid punishment. DePaulo and Jordan
1982, cited in Stouthamer-Loeber, 1986) speculate that lying to obtain a reward
comes at a latter stage of development than does lying to avoid punishment.

(g) Heighten Self-Presentation:
In contrast with lies to avoid a negative evaluation, lies to heighten self-presenta-
tion serve to present the perpetrator in a positive light (Petitclerc & Hervé 1999).
This may be similar to a faking good strategy, defined as a tendency to deny
symptoms. On the other hand, “faking bad” is defined as a strategy of endorsing
symptoms (Hare et al., 1989). Although initially it may appear that the same
argument could explain avoiding negative evaluation, this category serves to
make the offender appear more normal, not good or bad. There is a large litera-
ture of faking good and bad (Bagby, Rogers, Buis, & Kalemha, 1994; Paulhus,
Bruce, & Trapnell, 1995). Many studies have looked at the association between
psychopathy and these strategies (Hare et al., 1989). It appears when employing
many measures (MMPI, EPQ, MCMI, & BIDR) that psychopathy is associated
with a tendency to claim symptoms rather than deny them (Hare et al., 1989).

(h) Altruistic:
Altruistic lies are motivated by the perpetrator’s desire to protect another from
some harm (Ford, 1996). Typically these lies take the form of lying to shield the
feelings of the target, or to protect another from negative consequences.

(i) Carelessness
Unlike compulsive lying, which is may be due to an impulse control problem
(Ford, 1996) and therefore beyond the offender’s control, the careless liar is in-
deed in control of the lying behavior. Assessing the content of the lies, which is of
secondary importance to the offender, can make the distinction. Here the offender simply does not care to give the target truthful information. Petitclerc and Hervé (1999) call these ‘amotivational’ lies because the offender has no motivation to comply with the target’s desire to gain information.

(j) Duping Delight:
These lies are, quite simply, motivated by the pleasure of deceiving another. For this reason, Ekman (1985) coined the term ‘duping delight’. Like the careless lie, the content is of secondary importance. What is primary is the offender’s desire to prove his ability to deceive, and to take pleasure in outwitting and conning a target. The less gullible the target, the more challenging and exciting the successful deception will seem (Ekman & Frank, 1993). Some authors talk about lies that are motivated by a desire to obtain a sense of power (e.g., Ekman, 1997; Ford, 1996). The sense of power comes from possessing information that the other one does not have, or from misleading the other in order to cause him or her to make wrong decisions.

Petitclerc and Hervé’s (1999) ten motivational categories of deception were employed in the present study to investigate their associations to psychopathy. As this model was never validated, this investigation was the first attempt to do so with an adolescent population. For the purposes of the present study, Ekman’s (1997) definition of lies was adopted. This requires two criteria. First, deception must be deliberate. Second, the target must not be warned (explicitly or implicitly) of the deception.

As discussed above, once a lie has been discovered using the previous criteria, the current study was concerned with distinguishing which deceptive motivations are found in offenders across psychopathy scores. For numerous reasons, the examination of antisocial personality disorder (APD; American Psychological Association, 1994 [APA]) was excluded from the current investigation. Although APD is important when looking at a civil psychiatric setting (Hart, 2001) as the presence of the pathology is lower and therefore has more discriminate ability, it is of little diagnostic significance in many forensic settings where virtually everyone has a record of arrest (APA, 1994). Hart (2001) suggests that professionals should avoid overestimating the significance of antisocial behaviours in the assessment of personality disorders in a forensic population. This is the case as 50-75% of offenders typically receive a diagnosis of APD (Hare, 1996).

A more forensically relevant personality disorder diagnosis is psychopathy (Hare, 1991) in that it serves to differentiate the more problematic offenders from those with APD. Although PCL-R scores are significantly correlated with diagnoses of APD, the relationship is asymmetrical. That is, most psychopaths meet the criteria to receive a diagnosis of APD, however, most of the offenders who are diagnosed with APD are not psychopaths (Hare, 1996; Hare, Cox & Hart, 1989). For that reason, APD can be seen as synonymous with criminality and is not the focus here.
Psychopathy was also the focus of the current study as it is a clinical construct defined by a unique constellation of affective, interpersonal, and behavioral characteristics (Hare, 1991), including pathological lying, manipulativeness, and deceit (Cleckley, 1976; Hare, 1991, 1996). Given their talent and tendency to deceive, psychopaths may be viewed as natural liars or performers (Ekman, 1985; Ekman, 1992), indeed lying and deception are key clinical features of the psychopath (Hare, et al., 1989). In many cases the goals of a lie are to obtain money, prestige and power (Hare et al., 1989). The literature addressing lying among delinquents highlights another reasons that deception is important to address in these populations. In a review of seven studies (Stouthamer-Loeber, 1986) roughly speaking half of the youth exhibited problems with lying. These studies illustrate the need for more research delving into the association between deception and psychopathy in adolescents. Additionally, due to the link between deception and crime, it is fairly certain that psychopaths are deceptive; the question remains whether they employ different deceptive strategies than other antisocial persons.

Accordingly, it was expected that the psychopathic individual would be motivated to lie for four reasons using the current model (Petitclerc and Hervé, 1999): they would lie to heighten self-presentation; they would be motivated by duping delight; to avoid punishment; and they would lie to obtain a reward. No differences were predicted across the other categories. In addition, it was hypothesized that the psychopath would lie more frequently, compared with other offenders. These predictions were also made due to the results of previous studies with adult psychopaths (Petitclerc et al., 2000 & Spidel, 2002) which found that psychopaths lied significantly more than non-psychopaths for duping delight, to avoid punishment, to heighten self-presentation, and to obtain a reward. Although the results were of importance, they did not address adolescent offenders. As such, the current study used the same paradigm (Petitclerc & Hervé, 1999) and expanded the focus to young offenders.

2 METHODS

2.1 PARTICIPANTS

Seventy participants (53 male and 17 female, average age 15.8) were selected from a database of juvenile inmates who participated in research conducted in youth facilities in Burnaby, British Columbia. To be selected, they had to have an audio taped personality interview on file, along with file information. The personality interviews are based on the semi-structured interview protocol from the Psychopathy Checklist-Youth Version (PCL: YV; Forth, Kosson, & Hare, in press).
3 | MEASURES

3.1 | PSYCHOPATHY

Psychopathy assessments were completed from file and interview information using the Psychopathy Checklist-Youth Version (PCL:YV; Forth, Hart, & Hare, 1996). The PCL: YV is a modification of the Hare Psychopathy Checklist-Revised (PCL-R; Hare, 1991). The PCL: YV is comprised of the PCL-R’s 20 items, three item (17, 18, 20) adapted for the assessment of adolescents.

The use of the PCL: YV is supported by a growing body of research demonstrating its reliability and validity in adolescent offender populations (Brandt, Kennedy, Patrick, & Curtin, 1997; Forth, Hart & Hare, 1990; Gretton, 1998; Mailloux, Forth, & Kroner, 1997; McBride, 1998).

As with the PCL-R, the PCL: YV items are scored on a three-point scale from 0 to 2. The individual item scores are added to yield a total score, which can range from 0 to 40, with a higher score reflecting a closer match to the prototypical psychopath. The current study employed the PCL-R as a categorical measure of psychopathy. The sample was divided into High (H: n = 16), Medium (M: n = 38), and Low (L: n = 16) PCL: YV groups, using the adult version recommended cut-off of 30 for the High group and 20 for the Low group (see Hare, 1991). Two raters coded 8 randomly selected files. The Spearman-Brown inter-class correlation co-efficient was .96, p < .000 for a single measure and .97, p < .000 for the average of the two measures.

4 | PROCEDURE

4.1 | SOURCE OF DECEPTION

Offender-perpetrated deception was identified by file and interview review. An offender’s statement was deemed to be a lie if it was found to be inconsistent across information available. Trained research assistants followed a strict protocol and identified motivations to deceive on a three-point scale (i.e., yes, maybe, no). Each lie was categorized into one of the 10 types of motivations. Subsequently, more general judgments were made on how characteristic or pervasive particular deceptive motivations were for each offender on a 3-point scale, with 0 being not at all characteristic, 1 being characteristic in some circumstances, and 2 being very characteristic. This is the summary motive rating for the offender.

Two raters were trained until they reached reliability in terms of detecting the lies and categorizing their motives. Ten training tapes and files were used. To be considered reliable the raters were trained using training interviews and files until they accurately detected the lies and categorized their motives for ten cases consecutively. Kappa coefficients were used to assess interrater reliability on the
dichotomous scores for each pathway. The reliability of detection of individual lies comparing whether or not the raters detected the same lies was .905, \( p < .000 \). Both raters identified the lies 94% of the time. The Spearman-Brown inter-class correlation co-efficient was .96, \( p < .000 \) for the reliability of the classification of individual lies into motives. The Spearman-Brown inter-class correlation co-efficient was .98, \( p < .000 \) for the summary deception coding.

5 | RESULTS

Offenders who scored high on the PCL-R were compared via two-tailed t-tests for the summary motive ratings to offenders who scored below 20 on the PCL-R. The summary motive ratings were used as the overall presentation of the offender was considered the most important measure in understanding their motives. High and low psychopathy groups were compared across 10 motivational categories. The Bonferroni correction was used to control type I error. Therefore, to be considered significant, \( p \) had to be equal to or less than .005. Significant differences were found across three of the deception categories. Psychopaths in the high group, compared to the low group, were found to lie primarily in order to heighten self-presentation (\( t(1,30)=5.270, p<.000 \)). Those in the high psychopathy group also used deception for duping delight (\( t(1,30)=4.077, p<.003 \)) and to obtain a reward (\( t(1,30)=3.850, p<.001 \)) more often than those in the low group.

Table 1. Means and standard deviations (SD) for psychopaths versus non-psychopath offenders across significant motivations

<table>
<thead>
<tr>
<th>Motivation</th>
<th>Non-psychopath Mean</th>
<th>Non-psychopath SD</th>
<th>Psychopath Mean</th>
<th>Psychopath SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtain a Reward</td>
<td>.06</td>
<td>.25</td>
<td>.88</td>
<td>.81</td>
</tr>
<tr>
<td>Heighten Self-Presentation</td>
<td>.19</td>
<td>.40</td>
<td>1.38</td>
<td>.81</td>
</tr>
<tr>
<td>Duping Delight</td>
<td>.13</td>
<td>.50</td>
<td>1.06</td>
<td>.77</td>
</tr>
</tbody>
</table>

6. DISCUSSION

As was expected, those offenders scoring high on the PCL: YV significantly differed from low scorers with respect to motivations for deception. Psychopathic adolescents told more lies to heighten self-presentation, to obtain a reward, and perhaps most interesting, for sheer duping delight. These are similar to the results found with adult offenders (Petitclerc et al., 2000 & Spidel, 2002). The types of lies in which psychopaths engage, even from an early age, correspond with many of the personality characteristics that are so strongly associated with the clinical construct of psychopathy, described above.
Several other areas need to be examined in future research. The ability to deceive has been related to adjustment and general social skills by Wheldall & Alexander (as cited in Lewis, 1993); therefore, future studies should examine these differences across psychopathy and investigate whether any such differences remain consistent into adulthood. If, as we noted, certain motivations for deceptive behaviour are an early part of development, the potential exists for predicting antisocial behaviours; in studies that examined lies in general, lying at a young age moderately predicted later maladjustment (see Stouthamer-Loeber, 1986). Moreover, future research should investigate the variability in deceptive motivations across various types of psychopaths. Theory and research suggests that psychopathy might best be understood in terms of subtypes (Blackbrun & Coid, 1999; Hervé, 2002). Recent empirical work has identified three types of psychopaths - classic, macho, and manipulative, as well a subtype that appear to mimic psychopathy, most notably pseudopsychopaths (Hervé, 2002). It would be expected that as the classic and manipulative subtypes have similar interpersonal scores, they would therefore engage in deception for similar reasons. Although both are characteristically manipulative, the later is more likely to engage in deception than all other types, especially in regards to defrauding others. On the other hand, macho and pseudo have less interpersonal skills than the other two subtypes and may therefore be less likely to engage in certain types of deception that requires more verbal skills - the basis of manipulativeness. However, unlike idiopathic psychopaths, the pseudopsychopaths, presumably having some, although limited, emotional ties to others, are more likely to use deception to protect other people. As such it may be best when discussing types of deception to assess them across subtypes of psychopaths.

Several precautions need to be taken when interpreting these results. These data have implications only for convicted, young offenders. Accordingly, we are not in a position to make statements regarding other relevant deceiver characteristics. In addition, the current study did not look at gender differences, which may moderate the relationship between deceptive motivations and personality disorders. Gender differences have in fact been found in deceptive behaviour with girls being able to mask their internal feelings at an earlier age than boys (Lewis, 1993). Following Lewis et al.’s (as cited in Gervais, Tremblay & Desmarais-Gervais, 2000) hypothesis that girls possess more deceptive skill than boys, investigations into gender differences would be an important avenue for future investigation across personality-disordered adolescents.

A significant limitation of the current investigation is the sample size, which is too small to draw anything more than tentative conclusions from. Nevertheless, our findings are concordant with what would be expected theoretically given their presentation, and the findings adult psychopaths (Petitclerc et al., 2000 & Spidel, 2002) and as such are useful as preliminary conceptualizations of how deceptive motivations and personality disorders interact in juvenile offenders.
Notwithstanding the above limitations, the present findings have implications for assessing an offender’s credibility, if the likely deceptive motivations of a psychopath are understood and known, the interviewer can be on guard for these types of lies. Further investigation may give insight into which situations psychopathic offenders are likely to lie. Assessment procedures can be tailored accordingly to facilitate an interviewer’s ability to detect offender lies. For example, knowing that a psychopathic offender is likely to lie to heighten self-presentation in an interview can alert them to probe areas where s/he is describing situations that make her/him look better than normal. Therefore, it is important to be mindful of differing lie content when interviewing inmates, depending on the psychopathy score, knowing that the likelihood of certain lies arising varies as a function of psychopathy.

Although understanding the motivations of deception may aid in assessing the credibility of statements made by the offenders in various situations, it may also raise new questions and controversies. On one hand, knowing that a psychopathic offender may lie to dupe the target may cause the psychologist working with them to be more vigilant and wary of their claims of improvement (Kosson, Gacono & Bodholdt, 2000). On the other hand, offenders who are assessed as psychopathic, who may benefit from therapy, may be incorrectly assessed as not having profited due to being labeled psychopathic. Therefore, although the findings are of importance in terms of assessing offenders statements, care must be taken not to assume all statement of this nature made by psychopaths are false.

REFERENCES


Victims and offenders


INTRODUCTION

Instruments such as the Violence Risk Assessment Guide (VRAG, Webster, Harris, Rice, Cormier, & Quinsey, 1994) and the HCR-20 (Webster, Douglas, Eaves, & Hart, 1997) have been developed to assess violence recidivism risk. A major criticism of such guides is that they fail to distinguish between different types of violence. That is, underlying the use of these instruments is the assumption that all violent men are amenable to the same risk criteria merely because they are violent. This clear example of tautological logic is perhaps what leads some researchers to conclude that violence risk prediction is of less utility than the toss of a coin (see for example: Monahan, 1981).

This same criticism can be applied to previous domestic violence risk prediction research. Recently, forms of violence have begun to be disaggregated to form specific risk assessment measures or guides, such as the Spousal Assault Risk Assessment (SARA; Kropp, Hart, Webster, & Eaves, 1994, 1995, 1998) guide and the Sexual Violence Risk (SVR-20, Boer, Hart, Kropp, & Webster, 1997) guide. Measures such as the SARA have been developed based upon several static (fixed) factors that have been found to be associated with spousal violence recidivism. However, these same factors have been found by researchers such as Hanson and Bussiere (1998) and Quinsey, Harris, Rice and Cormier (1998) to predict both general and violent recidivism among a variety of different offenders (also see for example: Monahan & Steadman, 1994; Webster, Harris, Rice, Cormier, & Quinsey, 1994; Hall, 1987; Monahan, 1981), and there is little empirical indication that these factors are unique to the prediction of wife assault. In fact, Hilton, Grant and Rice (2001) have concluded, “that violent recidivism by
serious wife assaulters is well predicted by the same variables that predict violent recidivism among offenders in general."

Researchers such as Kropp and Hart (1997), on the other hand, argue that domestic violence prediction requires the consideration of some variables that differ from those used for general violence, and the failure of a large body of validated research on the correlates of psychopathy to apply to batterer samples lends support to this position. Psychopathy is a personality disorder marked by “a persistent disregard for social norms and conventions; impulsivity, unreliability, and irresponsibility; lack of empathy, remorse and emotional depth; and, failure to maintain enduring attachments to people, principles, or goals (Hare, 1991, p. 45).” It is assumed that psychopaths lack the requisite traits for inhibiting aggressive or antisocial behaviors such as empathy, guilt, fear of punishment and the ability to form close emotional bonds (Cleckley, 1976; Hare, 1991; McCord & McCord, 1964; Miller & Eisenberg, 1988). As such, it comes as no surprise that criminal psychopaths, in comparison to other criminals, commit a disproportionate number of both general and violent crimes; psychopaths are said to make up 15-30% of the total Canadian correctional population (Hare, 1991; Harris, Rice, & Cormier, 1991).

Overall, psychopaths are more likely to target strangers, to commit violent acts while intoxicated (Hare et al., 1988) and to have motives of an instrumental rather than a reactive nature when compared to non-psychopaths (Cornell et al., 1996). In fact, Cornell et al. (1996) found that this latter distinction allowed for the differentiation of psychopaths and non-psychopaths, with victimization often occurring for revenge rather than self-defense (Hart, 1998). But, how well does what we know about psychopathy apply to psychopathic batterers? Research seems to indicate that psychopathic non-batterers can present quite differently from psychopathic batterers, especially in terms of criminal behavior. For example, Williamson et al. (1987) found that while none of the murders perpetrated by psychopaths were against family members, 63% of the non-psychopaths’ murders were. The researchers comment that this finding does not come as a surprise, as familial murders are often committed in the “heat of passion”; such emotional volatility is usually not characteristic of the emotionally/physiologically-controlled psychopath. It would appear that having psychopathic traits also has treatment implications for batterers. For example, Dunford (2000) evaluated 850 men from San Diego who had completed treatment for wife assault. Using a version of the PCL-R, the 12-item PCL Screening Version (PCL-SV), Dunford found that men who failed in treatment were more than six times as likely to score above 12 (scores ranging from 0 - 24).

Research has found psychopathy to be an excellent universal predictor of violent recidivism (e.g., Salekin, Rogers, & Sewell, 1996) and yet, there is some argument as to whether the most dangerous batterers are psychopaths (Huss, Langhinrichsen-Rohling, & Scalora, as cited in Hilton, Harris, & Rice, 2001) or whether this group is more likely to have Borderline or other personality disor-
der features (Dutton, 1998). Another instance comes from the observation that suicide has been identified among various wife assault checklists as an indicator of risk (e.g., Goldsmith, 1990; Kropp, Hart, Webster, & Eaves, 1995), suicide being rare among psychopaths (despite their tendency towards making suicide threats) (Cleckley, 1964, p. 393). Another example comes from Dutton’s (1998) work suggesting a fear of abandonment among batterers that clearly, for the promiscuous and detached psychopath, is not of great consequence. Thus, both the construct of psychopathy (being an excellent universal predictor of violent recidivism; Salekin, Rogers, & Sewell, 1996), and its most commonly used measure, the Psychopathy Checklist-Revised (PCL-R; one of the best predictors of general violence risk; Hart & Hare, 1997), provide an excellent means to assess whether spousal violence risk is best assessed specifically or vis-à-vis a general violence risk measure. The current study examines the relative efficacy of applying measures of general violence risk, such as the Psychopathy Checklist-Revised (PCL-R; Hare, 1980, 1990) (discussed below), versus measures of specific violence risk, such as the Spousal Assault Risk Assessment Guide (SARA; Kropp, Hart, Webster, & Eaves, 1994, 1995, 1998) (also discussed below), to batterer populations.

2 | METHOD
2.1 | PARTICIPANTS

The files of 92 men with a history of domestic violence were chosen from a database of adult federal offenders at Matsqui and Kent correctional institutions in Vancouver, British Columbia, Canada. Only the files of incarcerated males who had a criminal record for domestic violence or for whom file information indicated a history of spouse abuse were selected. Approximately 49 percent of the sample was Caucasian; 26 percent was Aboriginal; 8 percent of the sample was classified as “other” (2 percent was Asian, 1 percent was Southeast Asian, 2 percent was Afro-Caribbean, 1 percent was Hispanic, and 1 percent was Caucasian-Aboriginal); and the remaining 17 percent of the sample was composed of individuals whose files failed to specify race. Participants ranged in age from 25 to 72, with a mean age of 41 years.

3 | MEASURES
3.1 | CRIMINAL HISTORY

All offenses listed on the subject’s criminal record were recorded.
3.2 | SPOUSAL VIOLENCE

The severity, aggressiveness and impulsivity of any codeable spousal assault incidents were rated on a three-point scale. An inter-rater reliability alpha coefficient of .90 was found for the coding of severity, .93 for aggression and .95 for impulsivity. For the severity of the incident, which refers to the amount of harm inflicted on the victim, a “1” indicated mild physical/verbal abuse (e.g. slapping, pushing, grabbing, shoving, threats etc.), a “2” indicated moderate physical abuse (e.g. kicking, biting, hitting with fists threats to use a knife or a gun, etc.) and a “3” indicated severe physical abuse (e.g. using a knife or a gun, assault requiring emergency medical assistance, etc.).

Aggressiveness was defined primarily in terms of extent to which the force used during the assault was proportionate to the amount required to achieve the end. The “end” or goal of the assault could be internal or external; an example of an internal “end” would be if the offender used the assault to gain compliance. Here a “1” referred to mildly aggressive incidents. If this intimate protested enough, it is likely that the assault would have stopped. A “2” was indicative of moderate aggressiveness. It is unlikely that protest would have stopped the assault, but the force used was proportionate to that required to achieve the end (the victim may need medical attention, but no hospital stay is required). A “3” was used to note severe aggression, where the force used was disproportionate to that required (e.g. the victim required emergency medical assistance and will likely require a stay in the hospital). Such an assault is unnecessarily brutal.

Coding impulsivity involved rating the degree to which the incident could have been foreseen or planned out. Again, this item was coded on a three-point scale, ranging from “1” (the incident was coldly planned out and executed) to “3” (No foresight or planning. The act was committed in the heat of the moment).

3.3 | THE PSYCHOPATHY CHECKLIST-REVISED

Hare (1980, 1991) developed a valid and reliable method of assessing psychopathy: the Psychopathy Checklist-Revised (PCL-R). The PCL-R is a 20-item scale, which is scored from correctional and psychiatric records and from a clinical interview. Each item is scored on a 3-point scale (ranging from 0-2), based upon the extent to which the item is characteristic of the individual. Thus, total scores on the PCL-R range from 0-40, with 30 typically being the cutoff for a diagnosis of psychopathy (For more detail on the PCL-R, please see the manual; Hare, 1991). The checklist has been found to possess both high construct validity (Harpur, Hare, & Hakstian, 1989; Newman & Kosson, 1986) and high reliability (Ogloff, Wong, & Greenwood, 1990; Kosson, Smith, & Newman, 1990; Smith & Newman, 1990; Wong, 1988) in a variety of contexts. Two factors have been identified as comprising psychopathy as a construct. Factor 1, derived from Cleckley (1976), is the interpersonal/affective dimension and concerns the indi-
individual’s core personality features. Factor 2 (identified by Harpur et al., 1989) is the behavioral component and refers to the extent to which the individual’s life is antisocial/prosocial and stable/unstable.

Although most files contained a PCL-R score, all files were re-scored by individuals trained in the administration and scoring of the instrument. It is important to note that PCL-R scores rated exclusively from file information, as they were here, tend to be several points below PCL-R scores rated from both file and interview data (Alterman, Cacciola, & Rutherford, 1993; Wong, 1988).

3.4 | THE SPOUSAL ASSAULT RISK ASSESSMENT (SARA) GUIDE

The SARA is a set of professional guidelines developed by Kropp, Hart, Webster, and Eaves (1994, 1995, 1998) for assessing spousal violence risk. The evaluator examines nine areas of the offender’s life through standardized measures of emotional/physical abuse and substance abuse; interviews with the accused and victims; and a review of collateral sources including victim statements, police reports, criminal records and psychological reports. The nine areas are as follows: Family of origin history; occupational and social history; relationship history; physical and mental health history; current mental status, assaultive/abusive behavior history; criminal history; current life-stressors; and, current support.

Use of the SARA requires that the presence or absence of 20 risk factors be evaluated. These items are scored on a 3-point scale, ranging from “0” (absent) to “2” (present). Ten of the factors refer to general violence risk, while the remaining ten specifically relate to spousal violence. In addition to coding for the presence of these twenty risk factors, evaluators must also note any additional (case-specific) risk factors. That is, they must note any particular markers that are considered “critical” to the individual’s risk. And finally, the evaluator is required to assess the individual’s global risk. This decision must be based on consideration of factors such as nature, severity, likelihood, frequency, and imminence of future violence (spousal or otherwise).

For the purposes of this study, only SARA summary ratings of risk to others and to partners were considered. Summary risk ratings are reported on a three-point scale, where a “1” indicates low risk, a “2” signifies moderate risk and a “3” denotes high risk. A rating of risk towards others was reported for 91% (84/92) of the sample, and a rating of risk towards intimates was reported for 34% (31/92) of the sample. Kappa coefficients for both variables were found to be .94, suggesting high inter-rater reliability.

Kropp and Hart (2000) analyzed a total of 2,681 SARA ratings among six adult male offender samples. The researchers found a high degree of batterer heterogeneity in terms of individual risk factors and of perceived overall risk. High
Victims and offenders

inter-rater reliability was observed for judgments of the presence of individual risk factors and for global risk. Structured analyses also suggest moderate levels of item homogeneity and internal consistency for the risk factors. Finally, in terms of other measures related to risk for general and violent criminal behavior, the SARA was found to have good convergent and divergent validity. The researchers conclude that the SARA distinguishes between both recidivistic and non-recidivistic spousal assaulters and between offenders with and without a history of spousal violence.

4 | RESULTS

Of the 92 files in the sample, 64 percent (59/92) included a PCL-R total score, 52 percent (48/92) included a Factor 2 score and 55 percent (51/92) included a Factor 1 score. Interclass correlations (using a 2-way random effects model for absolute groups) marking the reliability of the PCL-R coding were found to be .83 for Total scores, .72 for Factor 1 scores and .87 for Factor 2 scores.

Although no significant racial differences were observed, the sample’s PCL-R Factor 2 scores were shown to significantly differ in terms of age ($F = 4.27, p < .01$). Participants were separated into one of four age groups: Subjects in Group 1 ranged in age from 25-to-35 (n = 31); those aged 36-to-45 were placed in Group 2 (n = 29); Group 3 was composed of those men aged 46-to-55 (n = 13); and, Group 4 was made up of subjects aged 56 and above (n = 10). Tukey post-hoc analysis showed significant differences between the 25-35 (Group 1) and the 56 and above (Group 4) ($p < .01$) age groups and between the 36-45 (Group 2) and the 56 and above (Group 4) ($p < .05$) age groups, with higher scores being associated with the younger age groups. An effect size of .84 for the comparison of Group 1 and Group 4 was calculated. For the comparison of Group 2 and Group 4, an effect size of .63 was observed. The results suggest a moderate-to-strong diminishing effect of age on Factor 2 scores. This is an unexpected finding, as one would imagine that the older an individual is, the more extensive their criminal history is likely to be, criminal history being one of the variables considered in calculating Factor 2 scores.

Correlations were run between PCL-R Factor 1, Factor 2 and Total scores and number of domestic, simple and sexual assault charges. Surprisingly, correlations between number of domestic and simple assault charges and PCL-R scores failed to reach significance ($p > .05$). A significant negative correlation was observed, however, between Factor 2 scores and number of sexual assault charges ($r = -.35, p < .05$). This finding remains significant following bonferroni adjustments to the critical alpha level. Multiple regression analyses were performed to assess the extent to which PCL-R Factor 1, Factor 2 and Total scores could be used to predict number of sex offenses. A significant finding was observed ($p < .001$), with an $R$ of .60 a corresponding $R$ square of .36 and an adjusted $R$ square of .31.
Significant positive correlations were observed between imminent risk to others (as measured by the SARA) and both Total PCL-R ($r = .54, p < .05$; this latter correlation, however, failed to reach significance once bonferroni adjustments were conducted) and Factor 2 scores ($r = .58, p < .01$). All correlations between PCL-R scores and perceived risk to partner, however, failed to reach significance ($p > .05$). PCL-R scores were entered into a multiple regression analysis to assess the measure’s ability to predict SARA risk ratings. Although PCL-R scores failed to significantly predict risk to partner, they do seem to be a useful predictor of risk to others ($p < .05$), with a multiple $R$ of .65, an $R$ square of .43 and an adjusted $R$ square of .31.

Of the 92 files examined, 40 (approximately 43% of the sample) had sufficient information to code at least one spousal violence incident. Variables of particular interest were the severity, the aggressiveness (force) and the impulsivity associated with the most recent spousal assault incidents. A significant positive correlation was observed between impulsivity and Factor 2 ($r = .68, p < .01$) and Total scores ($r = .62, p < .01$), and a significant negative correlation was calculated for the relation between aggressiveness and Factor 2 scores ($r = -.60, p < .05$). These correlations remain significant after bonferroni adjustments have been made.

Regression analyses were performed to assess the extent to which simple assault charges and severity of spousal assault incidents could predict a diagnosis of psychopathy. Using the typical cutoff of 30 (a score greater than 30 being evidence of psychopathy), files for which PCL-R scores were available were divided into two groups: Those with psychopathy and those without. Of the 59 files for which the PCL-R was scored (64% of the total sample), 24 (approximately half the sample) scored above the cutoff while 25 scored below. The results support the hypothesis. That is, having a history of simple assault charges and of severe wife abuse appears to predict the presence psychopathy, with a multiple $R$ of .44, an $R$ square of .19 and an adjusted $R$ square of .15 ($p < .05$).

### DISCUSSION

As stated, some researchers (e.g., Hilton et al., 2001) have concluded that spousal violence risk is best predicted by the same variables that predict general violence recidivism. The PCL-R is considered to be one of the best predictors of violent recidivism (Hart & Hare, 1997). As such, following the logic of researchers such as Hilton et al. (2001), the PCL-R should predict intimate violence risk just as well as it predicts general violence risk and that specific violence risk measures such as the SARA have nothing unique to offer in comparison. The results of the current study fail to support this conclusion. Significant positive correlations were observed between imminent risk to others and PCL-R Factor 2 scores, but the same relationship was not found for risk to partner. Further, PCL-R scores functioned as significant predictors of risk to others but not risk to partner. These data agree with Kropp et al.’s (as cited in Dutton & Kropp, 2000) assertion that
when assessing spousal violence risk, it would be ill advised to use the PCL on its own. Thus, despite the fact that the PCL-R has proven to be a strong predictor of general violence risk (Hart & Hare, 1997), and the fact that it has shown to be a valid measure among batterer populations (Quinsey et al., 1998), results suggest that tools specific to predicting domestic violence risk (such as the SARA) may be better for assessing such risk than general violence risk measures (such as the PCL-R). Of course, without an outcome measure, further research is required.

Despite being one of the best predictors of violent recidivism (Hart & Hare, 1997), the PCL-R was not designed for this purpose; it was designed to assess for the presence of psychopathy (Hare, 1980, 1991). This does not by any means, however, render the measure useless when conducting batterer research. In fact, results of the current study suggest that the PCL-R could still be used to predict a batterer’s likely offense style. For example, a batterer with high Factor 2 or Total scores is less likely to have had sexual assault-related contact with the law and can be expected to assault their partners with less instrumentality and aggression than someone who scores lower on the PCL-R. Finally and perhaps more importantly, the psychopathic batterer is likely to be severe in his intimate assaults and to have a history of simple assault charges.

Clearly, any tool that improves the efficacy with which one can predict spousal violence risk would facilitate a reduction in the increasing number of male batterers coming in contact with the criminal justice system (Dutton & Kropp, 2000). The results of the current study lend support to the trend of developing specific violence risk assessment tools. That is, what we know about violence risk prediction, despite the fact that there is considerable overlap in the variables that seem to predict general and specific acts of violence, should be reframed in terms of various violent offender groups to develop specific risk measures, such as the SARA.

REFERENCES


Victims and offenders


THE SPOUSAL ASSAULT RISK ASSESSMENT (SARA) GUIDE AND THE PSYCHOPATHY CHECKLIST-REVISED (PCL-R)


Cognitive bias and judicial decisions

FRANCISCA FARIÑA, RAMÓN ARCE & MERCEDES NOVO

I | INTRODUCTION

In 1977, Ross classified the sources of bias that influence human judgements into motivational and cognitive. Motivational bias satisfies individual needs and desires whereas cognitive bias arises from limitations in human information processing. Motivational bias tends to lead to irrational judgements while cognitive bias involves, due to the limitations of information-processing strategies, judgements that systematically deviate from accepted norms or standards. According to Kruglanski and Ajzen (1983) three sources of cognitive bias are involved in attribution and prediction: information salience and availability, preconceived ideas or theories concerning people or events, and the phenomena of anchoring and perseverance.

1.1 | INFORMATION SALIENCE AND AVAILABILITY

The nexus between causal attribution and saliency, proposed by Heider (1958) and corroborated by Taylor and Fiske (1978), suggests that inferencing is based on information saliency and availability. Thus, the more available, probable, frequent, and salient the information is at the time of judgement making, the more it will guide causal inferencing (Plous, 1993).

On occasions, the judgement-maker makes inferences about a person’s behaviour on the basis of sampling bias. That is, when the sample of the information available is not representative of the entire population. Consequently, a judgement-maker may erroneously infer, bearing in mind atypical behaviour, personal dispositions.

[i] This research was supported in part by Programa Sectorial de Promoción General del Conocimiento (DGICYT). Code: PB96-0944.
Another potential source of bias derives from selective recall and perception. Taylor and Fiske (1975), having reviewed and integrated the literature, proposed that judgement-makers tend to look for a sole, sufficient, and salient explanation of behaviour, and that causal attributions are often modulated by salient stimuli. Even if all the relevant information is readily available to judgement-makers, they may selectively focus on those characteristics of the situation that are perceptually salient, and overestimate the importance of this information in later causal explanations.

A further source of bias related to information saliency is the “fundamental error of attribution”, also known as the Jones-Harris’ effect, which exaggerates the relevance of dispositional factors at the expense of situational or environmental factors in explaining behaviour (Jones & Harris, 1967). Some authors claim that the “fundamental error of attribution” can be explained in perceptual processing terms. That is, the person’s behaviour attracts and absorbs the judgement-maker’s attention, which will cause an overestimation of the causal importance of the person’s behaviour in comparison to other less salient factors (Fiske & Taylor, 1991). Another explanation lies in the internalisation of social norms, which implies that internal attributions tend to be more favourably evaluated than external ones. Both explanations have been experimentally corroborated (Sherman & Corty, 1984). This source of bias is not systematic and depends on the judgement-making demands, such as getting a general impression of the subject, comparing people or predicting behaviour (Papastamou, 1989). In cognitive terms, this implies that if a person’s behaviour is salient, causal inferencing linked to the person’s behaviour will be more available and prominent in the explanations (Moore et al., 1979). Moreover, implicit theories of people imply that dispositional causes will be overestimated in the explanation of behaviour. Though it has often been argued that the fundamental error of attribution is not as fundamental as previously thought, the observation of negative behaviour frequently over-attributes dispositional factors (i.e., attitudes, skills, personality traits) (e.g., Tetlock, 1985). The evaluation of negative behaviour is central to judicial judgement making. Thus, Ashworth (1984) reported that the vast majority of judges believed that the factors that inhibited most people from committing crime were their moral beliefs and the fear of social stigma.

Selective recall is also another source of cognitive bias, which is related to information saliency and availability. As Tversky and Kahneman (1973) have pointed out, a person uses the heuristic of availability when they evaluate the frequency or value of an event on the basis of examples that are readily accessible. Thus, in the legal context, judgements are influenced by the ease with which one can draw on examples. Nevertheless, selective recall does not imply that people cannot recover dissonant information, but rather that it is not used in the explanation of the events due to the task demands (Bekerian & Dennet, 1988; Arce, Fariña & Novo, 1997).
1.2 | PRECONCEIVED IDEAS OR THEORIES ABOUT PEOPLE AND EVENTS

Preconceptions guiding the use of information in the prediction and explanation of events incline people towards certain hypotheses and information, and predispose them to adopt intuitive ideas about an event or behavior at the expense of relevant information (e.g., Bar-Hill, 1980; Kelley, 1972). Preconceptions are said to arise from three sources: presumed covariation, representativeness, and causal theories.

The classical Asch (1946) studies on the formation of impressions (i.e., context effect, central traits) have revealed that intuitive comprehension of the relations between variables can exert an influence on our judgments. Kelley (1972) describes a pairing and grouping scheme to represent preconceived ideas related to the covariation between events. An example of the covariation between personality traits and behavior that can lead to bias and errors is the illusory correlation (Chapman & Chapman, 1969), referring to the erroneous belief that two unrelated variables are in fact related.

The heuristic of representativeness (Kahneman & Tversky, 1973) is another source of bias that rests on preconceived theories or ideas. Succinctly, when people have to evaluate the probability of a fact such as whether an object A belongs to a class B, they tend to use this heuristic. For example, “when A is highly representative of B, the probability that A originates from B is judged to be high. On the other hand, if A is not similar to B, the probability that A originates from B is judged to be low” (Tversky & Kahneman, 1974/1986 p. 39). Though it provides a quick solution, it entails bias such as: neglecting prior odds, disregarding the size of the sample, insufficient regressive judgement-making, underestimating the predictive value of evidence; or ignoring the fact that information stored in memory is not always reliable.

Causal theories underline the role of causal schemas in the formation and revision of beliefs (e.g., Tversky & Kahneman, 1980). The causal scheme presupposes a preconception about how two or more causes interact and the effect they produce (Kelley & Michela, 1980). That is, people may infer that several separate factors cause something to occur; the existence of one of these factors may lead one to imply the other(s). Kelley’s (1972) model describes two causal schemas: multiple causes and sufficient multiple causes. Whereas Kelley underlines the economic function of these types of schemes, Kahneman and Tversky (1973) emphasize the nature of them as a potential source of bias in judgement making.

1.3 | ANCHORAGE AND ADJUSTMENT

The last major source of cognitive bias stems from the anchorage and adjustment heuristic. Information saliency and availability or preconceived ideas can drive an
initial hypothesis, which may serve as an anchor (Tversky & Kahneman, 1974/1986) that restricts cognitive activity or serves as a cognitive set (Azjen, Dalto & Blyth, 1979) to guide the interpretation of the new information (e.g., Kaplan, 1982).

Judicial systems assume algorithmic strategies that one takes into account all the possibilities of the problem. Nevertheless, Saks and Kidd (1986) have pointed out that judicial decision-making is a good example where the probabilistic and uncertain nature of the task gives rise to the use of heuristics or simplification tools that reduce the complexity of the information required for decision making. This had led these authors to design a judicial decision-making model based on heuristics. Likewise, Fitzmaurice and Pease (1986) have proposed a model based on bias in judgement making. In line with these studies, the present work aims to evaluate if judicial decisions are mediated by heuristics strategies that lead to bias and error or by algorithmic ones (Kruglanski & Azjen, 1983). Thus, our aim is to search for cognitive bias in judges’ written judgements as well as identifying the likely cognitive activity on which they are based. In line with the scientific literature, motivational bias will be ignored in this paper since it responds to very irrational tendencies, and is not contemplated in judicial reasoning.

2 | METHOD

2.1 | PROTOCOLS

A total of 555 criminal judgements (in an inquisitorial system) were selected from the Appeal and High Criminal Courts of the Autonomous Community of Galicia, spanning the period from 1980 to 1995. As for the verdict, 457 (82.3%) were guilty, 93 (17.7%) not guilty, and 5 (.9%) of cases were not admitted for lack of evidence and were recodified as not guilty. Disregarding the latter, guilty judgements were significantly greater, \( \chi^2(1) = 240.9; p<.001 \). One hundred seventy-two judgements were High Court (31.0%) and 383 (69.0%) Appeal Court.

In terms of the case type, 139 (20.62%) were grievous bodily harm (GBH); 75 (11.12%) crime against property; 67 (9.94%) traffic offences; 41 (6.08%) drug and public health offences; 40 (5.93%) minor offences i.e., threats, resisting arrest; and 40 (5.93%) “fraud”. A glance at Table 1 shows the probability of guilty verdicts in relation to the nature of the offence.
Table 1

<table>
<thead>
<tr>
<th>Crime</th>
<th>Not guilty</th>
<th>Guilty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor offences</td>
<td>21.1%</td>
<td>78.9%</td>
</tr>
<tr>
<td>Drug and public health offences</td>
<td>21.95%</td>
<td>78.05%</td>
</tr>
<tr>
<td>Traffic offences</td>
<td>22%</td>
<td>78%</td>
</tr>
<tr>
<td>Grievous bodily harm (GBH)</td>
<td>27.41%</td>
<td>72.59%</td>
</tr>
<tr>
<td>Crimes against property</td>
<td>30.66%</td>
<td>69.34%</td>
</tr>
<tr>
<td>Fraud</td>
<td>40%</td>
<td>60%</td>
</tr>
</tbody>
</table>

2.2 | ANALYSIS OF PROTOCOLS

The protocols, that is the judgements', consist of two sections, the first referring to the facts, and the other to legal considerations. Both have been included in the analysis.

The first objective of the analysis of these protocols was to search for cognitive bias. Definitions and examples are presented as follows:

2.3 | INFORMATION SALIENCY AND AVAILABILITY

Though all of the relevant information may be available, the judgement-maker may only focus on some of the salient characteristics or data, and fail to consider other relevant information. An example of selective attention is, “...the victim was left with a “C” shaped scar that will be clearly visible if he were to go bald...”. Yet the scar was not really visible, and the defence’s allegations of self-defence were disregarded without any argumentation. Thus, the judgement-maker focuses on specific information neglecting other relevant evidence. A pattern of “fundamental error of attribution” is: “...Bearing in mind the dangerous means employed (a steel pitch fork), and the action of hitting Mr. R. in the area of the heart, confirms the intention of grievous bodily harm which excludes the lack of responsibility and any mitigating circumstances proposed by the defence...”. In this example, the judge is overestimating the personal disposition, disregarding situational factors indicative of self-defence. Moreover, selective recall was evaluated by referring to previous cases, in terms of the law of precedence.

[2] The term judgement refers to the judge’s or court’s written decision. In some countries the term sentence is used, whereas in others the term sentence is used only to refer to a guilty verdict. In this paper, the term judgement is used to refer to the judge’s or court’s written decisions whether it be a guilty or not guilty verdict.

[3] Selective recall and sample bias were not detected in any of the sentences.
2.4 | PRECONCEPTIONS

Preconceptions are preconceived ideas concerning the covariation of events, causal theories or representativeness; that is, when the judgement-maker has preconceived ideas of how things occur and how things normally occur. A causal example is: “the accused at the time of arrest for drunken driving, was unable to park the car near the curve, refused to be breathalysed and insisted that if he were to be breathalysed it should be done not in Lugo but in Castroverde. This together with the fact that the accused had ingested wine and gin-tonic before being stopped convinces the court that ...”. An example of representativeness is: “...the events occurred in the city of Lugo, but the accused is from Cambados, which supports the judge’s view...”. Finally, an example of a covariation model is as follows: “given that he is short sighted, and at the time he was not wearing his glasses, he did not realise what he was picking up from the counter...”.

2.5 | ANCHORAGE

Anchorage was defined as the prosecution’s request in terms of a judgement (Garrido & Herrero, 1997), or, in the cases of an appeal, the judge’s previous decision (Fitzmaurice & Pease, 1986). Anchorage was measured via the decision to incarcerate or the length of the sentence. Thus, anchorage was measured via the initial and direct estimates (Saks & Kidd, 1986; Wagenaar, 1995).

Since the aim was to evaluate possible differences in cognitive processing in the reconstruction of events, a system of categories was used that has proven to be useful and reliable in other studies (i.e., Fariña, Fraga and Arce, 2000). This system enables us to examine the underlying processes that influence judgement making. There are two category lists, one referring to general and the other to specific cognitive activity. The former deals with the number of words, general and specific thoughts, (the unit of analysis is the grammatical sentence, and it is specific when it is related to the case and general when it does not). Given that specific cognitive activity interacts with the content of the case in question, the two coders were previously asked to codify a list of categories obtained from other studies (i.e., Fariña, Fraga and Arce, 2000). Moreover, a procedure based on successive approximations was used to identify new categories related to the case in question. The list and description of the productive categories employed in the present study are as follows:

[4] Lugo was 5 km away, but Castroverde was over 30 km away.

[5] A valid categorical system requires the categories to be mutually exclusive (Weick, 1985). Thus, it was necessary to delimit perfectly the bias in the sample and the representativeness. The difference lies in the origin of the information. Representativeness is when the origin of the bias stems from the decision maker’s own ideas or theories in which case this would be an example of a sample bias derives from the evidence.

[6] Cambados is a city well known for being one of the centres for drug-dealing in Spain.
IDIOSYNCRATIC INFORMATION. An account of the number of references a judge makes to his/her internal state, cognitive processes and/or emotions.

DESCRIPTION OF INTERACTIONS. Total number of descriptions of interrelated actions and reactions.

REPRODUCTION OF CONVERSATIONS. Total number of virtual reproductions of expressions, certain manners of speaking or other people’s vocabulary.

CONTEXTUAL INCRUSTATION. Total number of embeddings related to the law of precedence and jurisprudence.

AMOUNT OF LEGAL DETAILS. An account of the number of legal references mentioned in the judgement.

CONTEXTUAL INFORMATION. Total number of references made to places, dates, time periods, etc.

ATTRIBUTIONS ON THE ACCUSED’S MENTAL STATE. An account of the references made by the judge to the mental state or motives of the accused.

ATTRIBUTIONS ON THE PLAINTIFF’S MENTAL STATE. An account of the references made by the judge to the mental state or motives of the plaintiff.

PHYSICAL CAUSAL RELATIONSHIPS. Total number of physical-causal events, when there is a presumed nexus between two physical events.

TEMPORAL CAUSAL RELATIONSHIP. Total number of temporal-causal relationships, when there is the existence of temporal continuity between two events.

NUMBER OF PRO-ACCUSED STATEMENTS.

NUMBER OF NEUTRAL STATEMENTS.

NUMBER OF STATEMENTS AGAINST THE ACCUSED.

NUMBER OF WORDS.

NUMBER OF GENERAL STATEMENTS. NOT RELATED WITH THE EVIDENCE.

NUMBER OF SPECIFIC STATEMENTS. RELATED WITH THE EVIDENCE.

An analysis of the internal consistency of the scales showed an alpha of crombach of .84 for the general cognitive activity and .77 for the specific cognitive activity.
3 | TRAINING OF ENCODERS

The two coders who participated in the study were trained, and the correlations and concordance index between codes served to contrast and correct any bias. Codings that did not coincide were discussed in order to homogenise the criteria. Both encoders had previous experience in other studies that had used the same coding system with similar categories (Arce et al., 1995).

4 | RELIABILITY

Each encoder analysed half of the protocols in search of cognitive biases and the categories that measure cognitive processing. One week after finishing the original encoding, 10% of the protocols were encoded again in order to determine the between- and within-encoder consistency. The consistency was calculated using the Kappa Statistic for the categorical variables and the correlation for the discrete ones7 (see Table 2).

Table 2. Within- and between-encoder consistency in “cognitive biases”

<table>
<thead>
<tr>
<th>Cognitive bias</th>
<th>between encoders 1-2</th>
<th>between encoders 2-1</th>
<th>within encoder 1</th>
<th>within encoder 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchorage</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>Saliency and availability</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>Preconceptions</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
</tbody>
</table>

Note: *p<.001

Having contrasted the encoder scores (see Tables 2 and 3), the measurements appear to be consistent within- and between-encoders for both cognitive activity, and detection of cognitive biases.

[7] We should bear in mind that this index is not accurate since it is not sensitive to the correspondence of the counts, thus the exact correspondence of the counts was verified. With this safeguard, Carrera and Fernández-Dols (1992) report that a correlation greater than .70 is reliable.
Table 3. Between- and within-consistency of cognitive processes

<table>
<thead>
<tr>
<th>Variables</th>
<th>r12</th>
<th>r21</th>
<th>r1</th>
<th>r2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts statements</td>
<td>.99*</td>
<td>.99*</td>
<td>1.00*</td>
<td>.99*</td>
</tr>
<tr>
<td>Amount of legal details</td>
<td>.99*</td>
<td>.98*</td>
<td>.99*</td>
<td>.99*</td>
</tr>
<tr>
<td>Attributed accused mental state</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
<td>.99*</td>
</tr>
<tr>
<td>Attributed victim mental state</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
<td>.99*</td>
</tr>
<tr>
<td>Contextual incrustation</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
<td>.99*</td>
</tr>
<tr>
<td>Contextual information</td>
<td>.96*</td>
<td>.94*</td>
<td>.99*</td>
<td>.99*</td>
</tr>
<tr>
<td>Description of interactions</td>
<td>.99*</td>
<td>.98*</td>
<td>1.00*</td>
<td>.99*</td>
</tr>
<tr>
<td>Idiosyncratic information</td>
<td>.99*</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>Neutral statements</td>
<td>.99*</td>
<td>1.00*</td>
<td>.98*</td>
<td>1.00*</td>
</tr>
<tr>
<td>Physical causal relations</td>
<td>.87*</td>
<td>.98*</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>Pro-accused statements</td>
<td>1.00*</td>
<td>.82*</td>
<td>.99*</td>
<td>.99*</td>
</tr>
<tr>
<td>Related statements</td>
<td>.99*</td>
<td>.99*</td>
<td>1.00*</td>
<td>.99*</td>
</tr>
<tr>
<td>Reproduction of conversations</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
<tr>
<td>Specific statements</td>
<td>.99*</td>
<td>.84*</td>
<td>.99*</td>
<td>.99*</td>
</tr>
<tr>
<td>Statements against the accused</td>
<td>1.00*</td>
<td>.99*</td>
<td>.93*</td>
<td>.982*</td>
</tr>
<tr>
<td>Temporal causal relations</td>
<td>.99*</td>
<td>.97*</td>
<td>.99*</td>
<td>.99*</td>
</tr>
<tr>
<td>Words</td>
<td>.84*</td>
<td>1.00*</td>
<td>1.00*</td>
<td>1.00*</td>
</tr>
</tbody>
</table>

Note: *p<.001

Furthermore, the encoders have been shown to be consistent in other contexts (Arce et al., 1995); thus the results can be defined as reliable (Wicker, 1975).

5 | DATA ANALYSIS

A multivariate analysis of variance (MANOVA) was carried out to assess cognitive activity associated to the presence vs. absence of the cognitive bias in sentencing. Among others, the reasons for preferring MANOVA to other tests were that it takes into account the intercorrelations among variables; keeps the overall level under control; gives univariate analysis (Stevens, 1986 p. 143). As is well known, the analysis of variance is a robust test, in particular with similar sized groups (large/small <1.5). Though many authors do not consider this to be of importance (Stevens, 1986), the absence of homogeneity of variance can lead to important deviations in the significance of the results. Thus, if the variability is greater in the small group, the F is liberal. In contrast, if the variance is greater in the large group, the statistic is conservative. In our study, some comparisons are made between two different sized groups. Consequently, as a safeguard, the variables were transformed using the square root to homogenise the variances.

[8] Nevertheless, the means presented in the text and tables are raw data.
Victims and offenders

(Dixon & Massey, 1983, pp. 373). As a second safeguard, the theoretical F of Box was used to confirm the correct acceptance or rejection of the hypothesis. Thus, if this were smaller than the empirical the alternative hypothesis was accepted, and vice versa (Palmer, 1996). The safeguards revealed no significant change in the results or in the regions of rejection or acceptance. In relation to multivariate tests the Pillia-Bartlett trace was used given that it is more robust to the effects of heterogeneity of the variance matrices (Olson, 1976).

6 | RESULTS

6.1 | CONTINGENCY OF COGNITIVE BIAS

Table 4 shows the frequencies of the presence/absence of cognitive biases in judgements. In the Spanish context, it appears that on the whole, the vast majority of judgements (74.95%) contained cognitive bias. Over half of the judicial decisions were based on anchorage (63%), about 16% on preconceived ideas or theories, and 9% on saliency and availability of the information.

Table 4. Contingency of cognitive biases

<table>
<thead>
<tr>
<th>Cognitive bias</th>
<th>Presence Frequency</th>
<th>Absence Percentage</th>
<th>Presence Frequency</th>
<th>Absence Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saliency and availability</td>
<td>50</td>
<td>9</td>
<td>505</td>
<td>90</td>
</tr>
<tr>
<td>Preconceptions</td>
<td>88</td>
<td>15.9</td>
<td>467</td>
<td>84.1</td>
</tr>
<tr>
<td>Anchorage</td>
<td>353</td>
<td>63.6</td>
<td>202</td>
<td>36.4</td>
</tr>
</tbody>
</table>

6.2 | COGNITIVE BIASES AND VERDICT

Anchorage tended to be systematically associated to guilty verdicts, $^2(i, n=550)=12.58; p<.001; phi = -.15$. That is, 87.4% of the guilty verdicts were linked to anchorage whereas decisions with no anchorage were observed in 75.6% of guilty verdicts. The relationship between anchorage and a guilty verdict lies at the very heart of the judicial process itself, since the process must commence with a firm accusation against the accused, which is the initial hypothesis of a guilty verdict that will serve as anchorage. When an initial hypothesis serves as anchorage i.e., as the point of departure on which to base the final estimates, it usually leads to bias or error of judgement. As Ross and Lepper (1980) have pointed out, the initial hypothesis perseveres in spite of the existence of information to the contrary.

[9] D.F.=1;n-k/k
In short, the findings suggest that this bias is used to subordinate more objective means of information processing such as normative inferencing models.

The source of bias “preconceived ideas or theories about people or events” tends to be associated with a significant reduction in the number of guilty verdicts, $\chi^2(1, n=550)= 4.88; p<.05; \phi = -.094$. It is worth noting that most of the decisions were free of bias based on preconceived ideas leading to 84.6% of guilty verdicts, which fell to 75% when the judgement-maker based his/her judgements on preconceived ideas. Initially, one would not expect a systematic tendency associated between this source of bias and the verdicts. The most reasonable explanation is the lower costs involved in a not-guilty verdict, that can be sustained with weak arguments derived from preconceived theories or ideas. Nevertheless, further research is required to determine the relationship between not-guilty verdicts and the judges’ preconceptions.

As for “information saliency and availability”, this cognitive bias was not systematically related to any particular judgement, $\chi^2(1, n=550)= .37; \text{ns}$.

6.3 | ANCHORAGE AND COGNITIVE ACTIVITY

In relation to the factor anchorage (presence vs. absence), the results exhibit a multivariate effect in general cognitive activity, $F_{\text{multivariate}}(3,55)=6.44; p<.001$. The univariate effects (see Table 5) show that in judgements without anchorage there is a greater number of words, and more reasoning related to the evidence in comparison with the judgements guided by the bias anchorage. In other words, the judgements without anchorage display more cognitive activity in order to explain the judgement reached. Thus, anchorage involves cognitive saving in motivating the judgement.

Table 5. Univariate effects in the dimension “general cognitive activity”

<table>
<thead>
<tr>
<th>Variable</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>eta2</th>
<th>M0</th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>General statements</td>
<td>.89</td>
<td>.01</td>
<td>.928</td>
<td>.000</td>
<td>2.48</td>
<td>2.56</td>
</tr>
<tr>
<td>Specific statements</td>
<td>182</td>
<td>12.98</td>
<td>.000</td>
<td>.023</td>
<td>48.03</td>
<td>36.1</td>
</tr>
<tr>
<td>Words</td>
<td>3328</td>
<td>6.26</td>
<td>.013</td>
<td>.012</td>
<td>753.95</td>
<td>595.04</td>
</tr>
</tbody>
</table>

Note: D.F. (1,553); $M_0=$ mean of the judgements without the cognitive bias “anchorage”; $M_1=$ mean of the judgements driven by “anchorage”.

Similar to general cognitive activity, the results show significant multivariate differences in specific cognitive activity mediated by the anchorage factor $F_{\text{multivariate}}(13,54)= 6.81; p<.001$. Likewise, at a univariate level, some variables
Victims and offenders appear to be mediated by this factor (see Table 6): physical and temporal causal relations, the number of legal details, neutral and pro-accused reasoning, description of interactions, contextual information, reproduction of conversations, and attributions to the mental state of the plaintiff. Thus, the absence of this bias is linked to more attributions to the mental state of the plaintiff, more descriptions of the interactions, legal details, contextual information, neutral propositions, pro-accused reasoning, to the establishment of more physical and temporal causal relations, and the reproduction of conversations. In other words, the absence of anchorage implies judgments which are more “driven to the facts” (contextual information, description of interactions, and reproduction of conversations); more “legally motivated” (legal details); and “causally guided” (temporal and physical causal relations). Furthermore, the absence of anchorage is closely linked to “attributions to the plaintiff’s mental state”, which have been identified as a tool to reject a guilty verdict on the basis of the plaintiff’s mental disorder.

### Table 6. Univariate Effects in the Dimension “Specific Cognitive Activity”

<table>
<thead>
<tr>
<th>Variable</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>eta²</th>
<th>M₀</th>
<th>M₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of legal details</td>
<td>1994.57</td>
<td>47.23</td>
<td>.000</td>
<td>.079</td>
<td>7.87</td>
<td>3.85</td>
</tr>
<tr>
<td>Attributed accused mental state</td>
<td>34.4</td>
<td>2.45</td>
<td>.118</td>
<td>.004</td>
<td>2.20</td>
<td>1.68</td>
</tr>
<tr>
<td>Attributed victim mental state</td>
<td>10.51</td>
<td>8.25</td>
<td>.004</td>
<td>.015</td>
<td>.54</td>
<td>.26</td>
</tr>
<tr>
<td>Contextual incrustation</td>
<td>.23</td>
<td>1.66</td>
<td>.198</td>
<td>.003</td>
<td>1.81</td>
<td>1.76</td>
</tr>
<tr>
<td>Contextual information</td>
<td>963.57</td>
<td>11.52</td>
<td>.001</td>
<td>.020</td>
<td>8.06</td>
<td>5.32</td>
</tr>
<tr>
<td>Description of interactions</td>
<td>981.26</td>
<td>7.42</td>
<td>.007</td>
<td>.013</td>
<td>9.40</td>
<td>6.63</td>
</tr>
<tr>
<td>Idiosyncratic information</td>
<td>.35</td>
<td>.084</td>
<td>.772</td>
<td>.000</td>
<td>.77</td>
<td>.82</td>
</tr>
<tr>
<td>Neutral statements</td>
<td>20154.38</td>
<td>10.19</td>
<td>.001</td>
<td>.018</td>
<td>33.54</td>
<td>21.01</td>
</tr>
<tr>
<td>Physical causal relations</td>
<td>11.16</td>
<td>5.95</td>
<td>.015</td>
<td>.011</td>
<td>.91</td>
<td>.61</td>
</tr>
<tr>
<td>Pro-accused statements</td>
<td>1284.15</td>
<td>21.56</td>
<td>.000</td>
<td>.038</td>
<td>5.49</td>
<td>2.32</td>
</tr>
<tr>
<td>Reproduction of conversations</td>
<td>128.27</td>
<td>6.72</td>
<td>.010</td>
<td>.012</td>
<td>2.06</td>
<td>1.06</td>
</tr>
<tr>
<td>Statements against the accused</td>
<td>138.89</td>
<td>.389</td>
<td>.534</td>
<td>.001</td>
<td>14.03</td>
<td>15.07</td>
</tr>
<tr>
<td>Temporal causal relations</td>
<td>1069.16</td>
<td>7.96</td>
<td>.003</td>
<td>.014</td>
<td>9.38</td>
<td>6.50</td>
</tr>
</tbody>
</table>

Note: D.F. (1, 553); M₀= mean of the judgements without “anchorage”; M₁= mean of the judgements driven by the cognitive bias “anchorage”.

The findings suggest that judgments based on anchorage rest on the process, evidence, categorization and reconstruction of the prosecution, or in the case of appeal courts the previous judge’s decision; whereas, the absence of anchorage requires a double process: rejecting anchorage and formulating a new judgement.
6.4 | SALIENCE AND AVAILABILITY

The bias “salience and availability” mediates significant multivariate differences in general cognitive activity, \( F_{\text{multivariate}} (3,55) = 13.65; p<.001 \). Univariate effects (see Table 7) reveal that judgements guided by the cognitive bias “information salience and availability” appear more closely connected to the evidence i.e., reasoning related to the case. Thus, it appears that there is no expected cognitive saving in the judgement.

Table 7. Univariate effects in the dimension “general cognitive activity”

<table>
<thead>
<tr>
<th>Variable</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>eta2</th>
<th>M0</th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>General statements</td>
<td>39.39</td>
<td>.36</td>
<td>.547</td>
<td>.000</td>
<td>2.45</td>
<td>3.38</td>
</tr>
<tr>
<td>Specific statements</td>
<td>8230.61</td>
<td>5.78</td>
<td>.017</td>
<td>.010</td>
<td>39.23</td>
<td>52.68</td>
</tr>
<tr>
<td>Words</td>
<td>57215.94</td>
<td>.11</td>
<td>.744</td>
<td>.001</td>
<td>650.41</td>
<td>685.88</td>
</tr>
</tbody>
</table>

Note: D.F.(1,553); M0= mean of the judgements without information “saliency and availability”; M1= mean of the judgements driven by information “saliency and availability”.

Likewise, significant multivariate differences were observed in specific cognitive activity modulated by the “salience and availability” factor, \( F_{\text{multivariate}} (13,54) = 4.21; p<.01 \). As for the univariate effects (see Table 8), differences were only observed in the variable pro-accused arguments. In other words, the presence of this bias tends to be associated more with pro-accused statements, which favour the accused. Nevertheless, this does not increase the number of not guilty verdicts as would be expected. Thus, the findings confirm that greater activity in favour of the accused is related to source of bias “salience and availability”, but this is not reflected in systematically biased not guilty judgements.

Table 8. Univariate Effects in the Dimension “Specific Cognitive Activity”

<table>
<thead>
<tr>
<th>Variable</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>eta2</th>
<th>M0</th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of legal details</td>
<td>9.64</td>
<td>.21</td>
<td>.647</td>
<td>.000</td>
<td>5.24</td>
<td>5.7</td>
</tr>
<tr>
<td>Attributed accused mental state</td>
<td>18.88</td>
<td>1.35</td>
<td>.247</td>
<td>.002</td>
<td>1.82</td>
<td>.46</td>
</tr>
<tr>
<td>Attributed victim mental state</td>
<td>.003</td>
<td>.00</td>
<td>.961</td>
<td>.000</td>
<td>.37</td>
<td>.36</td>
</tr>
<tr>
<td>Contextual incrustation</td>
<td>.000</td>
<td>.00</td>
<td>.972</td>
<td>.000</td>
<td>1.78</td>
<td>1.78</td>
</tr>
<tr>
<td>Contextual information</td>
<td>104.27</td>
<td>1.22</td>
<td>.269</td>
<td>.002</td>
<td>6.19</td>
<td>7.7</td>
</tr>
<tr>
<td>Description of interactions</td>
<td>169.59</td>
<td>1.27</td>
<td>.261</td>
<td>.002</td>
<td>7.47</td>
<td>.4</td>
</tr>
<tr>
<td>Idiosyncratic information</td>
<td>2.53</td>
<td>.61</td>
<td>.435</td>
<td>.001</td>
<td>.784</td>
<td>1.0</td>
</tr>
<tr>
<td>Neutral statements</td>
<td>2032.41</td>
<td>1.01</td>
<td>.315</td>
<td>.002</td>
<td>24.98</td>
<td>31.67</td>
</tr>
</tbody>
</table>
Victims and offenders

Physical causal relations 1.52 .8 .370 .001 .7 .88
Pro-accused statements 418.81 6.85 .009 .012 3.21 6.24
Reproduction of conversations 24.58 1.28 .259 .002 1.5 .76
Statements against the accused 300.66 .84 .360 .002 14.47 17.04
Temporal causal relations 29.18 .21 .644 .000 7.48 8.28

Note: D.F. (1,553); M0= mean of the judgements without information “saliency and availability”; M1= mean of the judgements driven by information “saliency and availability”.

6.5 | PRECONCEIVED IDEAS OR THEORIES OF PEOPLE OR EVENTS

The factor “preconceived ideas or theories” (presence vs. absence) mediates significant differences in general cognitive activity, F multivariate(3,55)=6.86; p<.001. The univariate effects (see Table 9) revealed that when this source of bias was detected in the formulation of judgements, judges resorted more to specific reasoning; that is to say, judgements were more connected to the evidence, and a greater number of words.

Table 9. Univariate effects in the dimension “General cognitive activity”

<table>
<thead>
<tr>
<th>Variable</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>eta2</th>
<th>M0</th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>General statements</td>
<td>18.56</td>
<td>.17</td>
<td>.679</td>
<td>.000</td>
<td>2.45</td>
<td>2.96</td>
</tr>
<tr>
<td>Specific statements</td>
<td>27776.95</td>
<td>19.99</td>
<td>.000</td>
<td>.035</td>
<td>37.37</td>
<td>56.74</td>
</tr>
<tr>
<td>Words</td>
<td>10365247.74</td>
<td>19.97</td>
<td>.000</td>
<td>.035</td>
<td>594.28</td>
<td>968.42</td>
</tr>
</tbody>
</table>

Note: D.F.(1,553); M0= mean of the judgements without preconceptions; M1= mean of the judgements driven by preconceptions.

Similarly, the preconception factor also modulates significant differences in specific cognitive activity, F multivariate(13,54)=5.38; p<.001. As for the univariate effects (see Table 10), these indicate that the reproduction of conversations, contextual information, references to the mental state of the plaintiff and the accused, and neutral reasoning are more frequent judgements driven by preconceptions. In contrast, preconceived written judgements contained fewer contextual incrustations. In other words, they were more driven “to the facts” i.e., involving more conversation and contextual information. These judgements were not “driven to the verdict” (with more neutral reasoning, with no significant reasoning either in favour or against the accused). On the other hand, preconceived judgements were based more on inferences, without the support of expert evidence, about the mental state of the accused and plaintiff. Moreover, preconceived jud-
gements are not as "legally driven" since they are grounded on weak jurisprudence i.e., with less contextual incrustation.

Table 10. Univariate effects in the dimension “specific cognitive activity”

<table>
<thead>
<tr>
<th>Variable</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>eta2</th>
<th>M0</th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of legal details</td>
<td>112.49</td>
<td>2.47</td>
<td>.117</td>
<td>.004</td>
<td>5.09</td>
<td>6.32</td>
</tr>
<tr>
<td>Attributed accused mental state</td>
<td>88.82</td>
<td>6.38</td>
<td>.012</td>
<td>.011</td>
<td>1.7</td>
<td>2.8</td>
</tr>
<tr>
<td>Attributed victim mental state</td>
<td>51.34</td>
<td>42.78</td>
<td>.000</td>
<td>.072</td>
<td>.24</td>
<td>1.07</td>
</tr>
<tr>
<td>Contextual incrustation</td>
<td>1.05</td>
<td>6.2</td>
<td>.013</td>
<td>.011</td>
<td>1.8</td>
<td>1.68</td>
</tr>
<tr>
<td>Contextual information</td>
<td>1449.54</td>
<td>17.51</td>
<td>.000</td>
<td>.031</td>
<td>5.62</td>
<td>10.05</td>
</tr>
<tr>
<td>Description of interactions</td>
<td>313.64</td>
<td>2.35</td>
<td>.126</td>
<td>.004</td>
<td>7.32</td>
<td>9.38</td>
</tr>
<tr>
<td>Idiosyncratic information</td>
<td>2.99</td>
<td>.72</td>
<td>.396</td>
<td>.001</td>
<td>.84</td>
<td>.64</td>
</tr>
<tr>
<td>Neutral statements</td>
<td>31659.36</td>
<td>16.18</td>
<td>.000</td>
<td>.028</td>
<td>22.3</td>
<td>42.98</td>
</tr>
<tr>
<td>Physical causal relations</td>
<td>1.57</td>
<td>.83</td>
<td>.363</td>
<td>.001</td>
<td>.74</td>
<td>.59</td>
</tr>
<tr>
<td>Pro-accused statements</td>
<td>48.33</td>
<td>.78</td>
<td>.377</td>
<td>.001</td>
<td>3.35</td>
<td>4.16</td>
</tr>
<tr>
<td>Reproduction of conversations</td>
<td>100.5</td>
<td>5.25</td>
<td>.022</td>
<td>.009</td>
<td>1.24</td>
<td>2.41</td>
</tr>
<tr>
<td>Statements against the accused</td>
<td>370.71</td>
<td>1.03</td>
<td>.310</td>
<td>.002</td>
<td>15.06</td>
<td>12.82</td>
</tr>
<tr>
<td>Temporal causal relations</td>
<td>161.87</td>
<td>1.19</td>
<td>.276</td>
<td>.002</td>
<td>7.32</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Note: D.F. (1,553); M0= mean of the judgements without preconceptions; M1= mean of the judgements driven by preconceptions.

7 | DISCUSSION

We should bear in mind that the results may not reflect the full extent of cognitive bias due to the possible limitations of our detection instruments. For example, content analysis of the judgements may not reveal the full impact of bias mediated by information saliency and availability. Furthermore, the results cannot be generalised to cover other legal systems, since they are derived from an inquisitorial system, with a specific jurisprudence and judgement-making context. Moreover, bias inferencing need not necessarily entail the likelihood of error even if inferencing procedures are not adequate (Kruglanski & Ajzen, 1983).

Taking into account these observations, our results reveal that:

a) Reasoning. Though judgement making is presumed to be based on formal reasoning, our results show that most legal judgements (74.95%) rest on informal reasoning i.e., bias. The main source of bias was anchorage, which mediated more than half of the judgements (63.6%).

b) Cognitive saving. Whereas as anchorage was found to be an important cognitive saving strategy for judgement making, information saliency and availability, and preconceptions were not. In short, the latter strategies require
inferencing though this does not imply that they are not cost-saving strategies in other spheres of the judicial process.

c) Verdicts. As expected, given the origins of anchorage in judicial proceedings, which are linked to the prosecution’s plea of guilt or to previous judgements sent to appeal courts, verdicts tend to be associated with a guilty outcome. In contrast, judges rely on preconceptions to reach a not guilty verdict. Finally, information saliency and availability have no relation with the verdict reached.

d) Information processing. In information processing, anchorage induces judges to exclude neutral information and information in favour of the accused which increases the incidence of guilty verdicts. Thus, anchorage leads to “information-exclusion processes” whereby both versions of the evidence are not considered before reaching a judgement. Preconceptions were linked to “information-integration processes” since judges equally process the information either in favour or against the accused, giving greater priority to neutral reasoning. In other words, both versions of the evidence are integrated. In comparison, judgements mediated by information saliency and availability tend to generate more reasoning in favour of the accused i.e., “pro-accused processing”, though this does not result in not guilty verdicts as would be expected.

e) Evidence. Anchorage enables a judge to arrive at decision without justifying it. Preconceptions, however, require a greater factual nexus for inferencing. As for bias derived from information saliency and availability the results were as expected i.e., related to the evidence.

f) Causal nexus. The reconstruction of events was less “causally guided” (both physically and temporally) in the decisions mediated by anchorage.

b) Ruling. The judgements associated to anchorage rested less on legal aspects (legal details, etc). As for the judgements based on the judge’s preconceptions, these referred less to the law of precedence.

In conclusion, judicial decisions are based to a large extent on informal reasoning, which is indicative of bias, as opposed to expected formal reasoning in which information is correctly treated (see Kruglanski & Azjen, 1983 for a review and discussion). Given that one of the anomalous sources of informal reasoning is metacognitive deficit, Perkins (1989) suggests that the solution to counter these sources of bias is to raise the judges’ awareness of possible sources of bias, in order to deal with metacognitive deficits in informal reasoning by ensuring greater objectivity in decision-making.

REFERENCES


Victims and offenders


Heider, F. (1958). The psychology of interpersonal relations. New York: Willey and Sons.


Victims and offenders


