Offending and Attachment: The Relationship between Interpersonal Awareness and Offending in a Prison Population with Psychiatric Disorder

Alice Levinson and Peter Fonagy
UCL (University College London), Gower Street, London, WC1E 6BT

ABSTRACT
This study aimed to test the predictions that, once psychiatric disorders are controlled for, offenders are more likely to be dismissive in their attachment patterns than are controls, and that the capacity of forensic patients to reflect on mental states of self and other is critically impaired, limiting their capacity to empathize and making them more liable to offend. Twenty-two prisoners were compared with 22 personality disordered patients without an offending history, and 22 normal controls. The Adult Attachment Interview was employed to examine early childhood trauma and attachment patterns; and the Reflective Function (RF) Scale was applied to measure capacity to reflect on mental states. The prisoners had experienced more abuse and neglect than the patients, yet were more likely to be coded resolved to their abuse on the AAI. As predicted, prisoners were more likely to be dismissive in their attachment patterns, and the prisoners’ RF was more impaired than that of the patients. Violent offenders showed the greatest deficits in RF. We suggest that prisoners’ developmental path of psychopathology is characterized by a disavowal of attachment-related experiences and of the capacity to think about them, in partial response to severe childhood trauma. The impairment of RF removes a critical barrier that might normally inhibit offending, leaving them more liable to act, especially in violent ways.

Introduction
Over the last decades we have learnt that the risk factors associated with antisocial behaviour in general and violence in particular are evident from relatively early childhood. The findings from reviews (Farrington, 2003; Loeber, Green, & Lahey, 2003; Rutter, Giller, & Hagell, 1998) and from some of the best-known longitudinal studies, including the Cambridge Study (Farrington, 1995), the Pittsburgh studies (Loebër, Stouthamer-Loebër, Farrington, Lahey, Keenan, & White, 2002), and the Dunedin Study (Moffitt, Caspi, Harrington, & Milne, 2002) have been fairly consistent, even if relatively small in magnitude. Repeatedly reported personality and temperament individual risk factors include: (1) uncontrolled temperament observed at age 3 leading to adult aggression (e.g., Caspi, 2000), (2) impulsive traits at age 8–10 associated with adult offending (e.g., Farrington, 1995), (3) hyperactive traits at 13 linked to adult violence (e.g., Klinteberg, Andersson, Magnusson, & Stattin, 1993), (4) callous traits at 7–12 predicting antisocial personality disorder in maturity (e.g., Loeber, Burke, & Lahey, 2002), and (5) low IQ and poor academic achievement predicting being arrested and charged (Moffitt, 1993). The risk of an adult antisocial personality disorder diagnosis has been shown to be increased by a range of childhood psychiatric conditions: (1) major depression at age 14 (e.g., Kasen, Cohen, Skodol, Johnson, Smailes, & Brook, 2001), (2)
oppositional defiant disorder diagnosed at 7–12 (e.g., Loeber, Burke, et al., 2002), (3) conduct disorder diagnosed at 9–16 (e.g., Harrington, Fudge, Rutter, Pickles, & Hill, 1991), and substance abuse observed at 7–12 (e.g., Loeber, Burke, et al., 2002).

Particular features of parenting and parents have also been good predictors of later offending and violence, including: (1) antisocial diagnosis and/or behaviour in the parent when the child is 8–10 (e.g., Smith & Farrington, 2004), (2) poor supervision of 8-year-olds (e.g., Farrington, 1995), (3) abuse when the child is under 12 (e.g., Widom, 1989), and (3) early or late exposure to domestic violence, which appears to predict not just violence but also violence to the individual’s own child (e.g., Moffitt & Caspi, 2003). In addition, of course, wider social factors have also been shown to play a part, including obvious risk factors such as association in adolescence with a deviant peer group (e.g., Lipsey & Derzon, 1998) and being in a high-delinquency school (e.g., Farrington, 1995).

Knowledge of risk factors on its own is of limited value. Risk factors with clear causal significance that are open to modification as part of prevention have the greatest practical relevance. Unmodifiable risk factors (e.g., gender) or modifiable risk factors that antedate the problem of violence but are not part of a causal process (e.g., peer delinquency) are of limited value. For example, peer delinquency may be as much a consequence of a delinquent and violent pre-disposition as its cause (Farrington, Loeber, Yin, & Anderson, 2002). Broadly speaking, the greater the number of risk domains entailed in the history of a particular case, the higher the risk of violent antisocial behaviour (Loeber, Stouthamer-Loeber, et al., 2002). Violence is the product of a chain of events over the course of a child’s development, where risks accumulate and reinforce each other (Maughan & Rutter, 2001). So, for example, low IQ places a child at increased risk of experiencing problems at school, which in turn can create major problem behaviours that lead to exclusion, and failing to graduate can lead to employment problems that in turn increase the risk of persistent antisocial behaviour. This is the argument for ensuring that violence-prevention programs must simultaneously target multiple risk factors. There is further valuable epidemiological information to be gained from understanding why certain factors appear to moderate the impact of risk factors. These characteristics, usually termed protective factors (Garmezy & Masten, 1994), appear to interrupt the causal chain of risk. For example, in the case of risk factors for violence, we know that characteristics such as shyness and inhibition, intelligence, a close relationship with at least one adult, good school or sporting achievements, and non-antisocial peers can positively moderate the impact of risk factors (Losel & Bender, 2003).

The new data from the clustering of developmental trajectories has brought a shift of emphasis to the developmental understanding of violence (Cote, Tremblay, Nagin, Zoccolillo, & Vitaro, 2002; Moffitt et al., 2002; Nagin & Tremblay, 2001; Shaw, Gilliom, Ingoldsby, & Nagin, 2002). The simultaneous analysis of the 6 samples by Broidy et al. (Broidy, Nagin, Tremblay, Bates, Brame, Dodge, et al., 2003) led the authors to conclude that there appears to be a continuity in problem behaviour from childhood to adolescence and that such continuity is especially acute when early problem behaviour takes the form of physical aggression. Chronic physical aggression during the elementary school years seems specifically to increase the risk for continued physical violence, as well as other non-violent forms of delinquency during adolescence. This was true, however, only for boys, because the results indicate no clear linkage between childhood physical aggression and adolescent offending among female samples. Historically, models of aggression have focused on how
human aggression is acquired through learning, rehearsal, and reinforcement of aggression-related knowledge structures (Anderson & Bushman, 2002). Yet aggression appears to be there as a problem from early childhood, arguably from toddlerhood, and perhaps from birth. Violence ultimately signals the failure of normal developmental processes to deal with something that occurs naturally. The key concerns of prevention are the individual, behavioural, family, and wider societal characteristics (including custodial services) of the individuals who do not desist from aggression during childhood.

At this juncture we should perhaps acknowledge Freud (1920/1961; 1930/1961) and classical psychoanalytical views that have consistently suggested—in line with modern developmental data—that social experience is there to tame a destructiveness that is inherent in humanity. Psychoanalysts expect to find violence in all individuals since it is a fundamental destructive urge, or, in Freud's words, an "independent aggressive instinct" (Freud, 1930/1961). Biological predisposition and social influence do not create destructiveness, but rather compromise the social processes that normally serve to regulate and tame it. Not that aggression always shows the failure of some system. We shall argue that the environment can spectacularly fail to provide the infant with the means to regulate its destructive potential. Violence may be the individual's attempt to tackle a damaging environment and as such can be "a sign of life" as Winnicott (1975, p. 85) called it, a sign of our struggles to carry on as living beings under intolerable conditions. The innate aggression theory must take proper account of the existence of positive, survival-oriented aggression and also of aggression that is a genuine protest against very considerable hardships in life.

An association between attachment and offending has also been long assumed. Disruptive family backgrounds including childhood separation and trauma are common in the history of offenders and delinquents (Bowlby, 1944, 1988; Lewis, 1989). Bowlby (1944) suggested in a study of a sample of offenders that the disruption of early attachment bonds may lead to adult attachment disorders—especially of an affectionless kind. Attachment theory provides a model to explain the links between emotional deprivation in childhood and the development of offending. The theory suggests that deprivation disrupts early attachment relationships, causes children to seek self-protection by avoiding or dismissing attachment relationships, and thus they do not have a way of forestalling the emergence of delinquency in the context of other personal and environmental risk factors (Rutter et al., 1998). The avoidance of attachment relationships may be a vulnerability factor that leads to a reduction of emotional commitment to social institutions (Hirschi, 1969) and may be directly involved in the aetiology of conduct disorder (Greenberg, Speltz, & DeKlyen, 1993), which is a risk factor for offending. While the point prevalence of dismissing attachment is common (20–30%) (van IJzendoorn, 1995), offending is relatively rare. Thus dismissive attachment on its own can no more provide an adequate model of offending behaviour than personality disorder and psychiatric illness. While neither dismissing attachment nor psychiatric illness and/or personality disorder can account for offending behaviour, it is possible that the two factors interact to increase the risk of offending.

We know of only one study using the Adult Attachment Interview (AAI) coding systems

1 In the text and Refs, van is sometimes lowercase, sometimes uppercase. But style in text and in Refs. Is parallel, e.g., here and in Refs it's lowercase. Shall we leave, as long as it's parallel? See next para where it's uppercase. It should be consistently lowercase
(George, Kaplan, & Main, 1985; Main & Goldwyn, 1991) reporting the relationship of attachment to offending. Van IJzendoorn and colleagues (Dozier, Stovall, & Albus, 1999; Van IJzendoorn, Feldbrugge, Derks, de Ruiter, Verhagen, Philipse, et al., 1997) studied 40 offenders in a forensic setting but could not differentiate the offenders from a psychiatrically disordered group. As psychopathology in general and personality disorder in particular are known to be associated with abnormal distribution of attachment patterns (Fonagy, Leigh, Steele, Steele, Kennedy, Mattoon, et al., 1996), and as psychiatric disorders, particularly personality disorders, are common in prison groups (see above), we felt that the distinctive character, if any, of attachment in offenders could be adequately examined only if psychiatric diagnoses were controlled for by matching cases and controls. In line with Bowlby’s hypothesis, we expected to find that when samples were matched for psychiatric and personality disorders, a higher proportion of prisoners’ AAI narratives would be classified as dismissive than controls.

We further hypothesized that a deficit in mentalizing (Frith & Frith, 1999) might be a critical mediating mechanism between insecure dismissing attachment and offending behaviour (Fonagy, Target, Steele, & Steele, 1997). Lack of a robust capacity to envision mental states in others might remove a critical barrier that normally inhibits behaviour that impinges on the rights of others, making such individuals more liable to cause harm. A somewhat retarded mentalizing capacity in insecurely attached children was demonstrated in a prospective study of 5-year-olds followed from birth (Fonagy, 1997). To investigate the relationship between mentalizing and offending, we have operationalized the former concept in terms of the construct of the Reflective Function scale (RF), described in Fonagy et al. (Fonagy, Target, Steele, & Steele, 1998). We hypothesized that offenders would score lower on this measure than matched controls.

**Method**

**Sample**

The prison group was selected by taking consecutive psychiatric admissions to the Health Care Centre (HCC) of a prison. The prison was a high-security and local remand prison, for male prisoners. The prisoners were referred to the HCC from reception when they entered the prison, for medical and/or psychiatric assessment and treatment. Inclusion criteria were (1) age range 20–40; (2) English as first language, to maximize the accuracy of the AAI coding; (3) a current, or if on remand a past, history of at least two convictions with at least one custodial sentence; and (4) at least one diagnosable Axis I or Axis II disorder, to permit matching of cases. The exclusion criteria were: (1) schizophrenia, (2) delusional psychoses, (3) organic brain disorder, and (4) an IQ of less than 80. Thirty-two subjects were referred during the study period. Eight did not meet the criteria (the majority on the basis of psychotic illness or organic brain disorders), and 2 did not agree to participate, leaving 22 male subjects in the prison sample. Remand and sentenced prisoners were included in equal numbers. The index offences ranged in severity and were divided for comparison into 2 sub-samples: less violent offences against property ($n = 10$) and more violent offences against persons ($n = 12$). The less violent offences against property included theft ($n = 1$), handling stolen goods ($n = 1$), burglary ($n = 4$), damage to property ($n = 1$), deception ($n = 1$), and drug supplying and importation ($n = 2$). The more violent offences against people included murder ($n = 1$), armed robbery ($n = 2$), malicious wounding/wounding with intent ($n = 2$), grievous bodily harm ($n = 1$), rape ($n = 2$), indecent assault ($n = 1$), and drink and driving ($n = 3$). The 3 drink-and-
driving offences were included in the violent group, because of the lack of concern expressed for others, and they each had coexisting current violent offences—one assault to the arresting police officer, the other two to attachment figures—as well as past violent offences.

The demographic and clinical characteristics of the sample are shown in Table 1. Social class was established on the basis of the UK Registrar General’s Classification of Occupations (General Register Office, 1966) according to occupation before imprisonment or, if unemployed, by father’s occupation. Nineteen subjects were white, 1 second-generation African Caribbean, 1 mixed-race African Caribbean and Canadian, and 1 African.

All subjects had at least one Axis I disorder, 80% having three or more. They all had at least one Axis II disorder, 50% meeting DSM–IV criteria for BPD.

The personality disorder comparison (PD) group, consisting of 22 personality-disordered patients, was recruited from an inpatient psychotherapy program and a community-based study of personality disorder. The selection criteria were identical for the prison group, except that individuals with histories of offending and/or convictions were excluded. Matching was on the basis of gender, age, IQ, social class, and approximate diagnosis on Axis I and Axis II. The mean overall number of diagnoses was somewhat greater in the prison group, but the Global Assessment of Functioning (GAF) scores were somewhat lower in the patient group. Matching was quite successful, given the very large number of comorbid diagnoses in both groups. Significant group differences were reached in only one diagnostic category. There was somewhat more anxiety in the clinical comparison group ($p < 0.05$) and slightly more cluster B personality disorders in the prison group (see Table 1).

The medical controls consisted of 22 subjects selected from a hospital medical department. They were a mixture of inpatients and outpatients from two medical units of a central London teaching hospital screened for psychiatric disorder using the GHQ (Goldberg & Williams, 1988), and all those above the clinical cut-off were excluded. They were matched for demographic criteria of gender, age, social class, and IQ.

Procedure
A psychiatrist briefly interviewed all prison subjects referred to the HCC when they entered the HCC, to establish initial inclusion and exclusion criteria. All the other interviews were conducted by the research psychiatrist (AL) one week later, to allow subjects some time to withdraw from drugs and to adjust to imprisonment. The AAI was then administered to all subjects and was audio-recorded. Mostly in the same interview session, the structured psychiatric assessments were administered to establish the range and intensity of psychiatric symptoms, personality disorder, and overall level of functioning. Finally, tests of reading ability and self-report instruments were administered, and a record of index and past offences was collected from the Home Office files for each prisoner. The PD and medical control groups were assessed, following the same sequence of evaluations, with the exceptions that the absence of a history of offending was established during initial screening, and the hospital comparison group was not administered structured psychiatric assessments.

Measures
1. The Adult Attachment Interview (AAI) (George et al., 1985). The AAI is a semi-structured interview used to obtain classifications of attachment representations, which are generally considered to correspond to Bowlby’s notion of internal working models. It consists of a series of questions and probes designed to elicit as full a narrative as possible about the subject’s
childhood attachment experiences. Experiences of loss, trauma (physical and sexual abuse), separation, and rejection are directly questioned, as well as the subject's evaluation of the effect of these experiences on their current functioning and adult personality. A description of their caregivers is asked for, followed by memories to illustrate these descriptions with the aim of eliciting information about the individual's current internal representations of childhood attachment figures and experiences.

The interview is audio-recorded and transcribed verbatim. The transcripts are rated on scales concerning experience (love, neglect, rejection, and pressure by caregivers) and states of mind (anger, idealization, derogation, and coherence of the narrative), as the latter has been shown to be the most powerful predictor of the attachment security of infants in trans-generational studies (van IJzendoorn, 1995). The AAI Scales are 1 to 9-point scales, with manualized definitions for each point.

The Main and Goldwyn coding system (Main & Goldwyn, 1994) distinguishes three major categories of interviews related to these scales. Secure transcripts, designated F for free, are coherent and balanced evaluation of past attachment experiences; insecure-dismissive transcripts (Ds) are characterized by an attempt to limit the import of attachment relationships by means of cut-off statements, lack of recall, idealization, and/or derogation and considering the self as unaffected by negative attachment experiences; and insecure-preoccupied transcripts (E) are characterized by passive, fearful, or angrily preoccupied and entangled narratives. In addition to the three main categories, subtypes of F, Ds, and E are also classified. For example, Ds1 is at the most severe end of the dismissive category, compared to Ds2, where there is some devaluation of attachment alongside some apparent continuing capacity for it, and Ds3 where a restriction of feeling is manifest in the narrative.

An Unresolved (Us) classification may be assigned to transcripts, superimposed upon any of the other classifications, if there is a failure fully to mourn a significant loss (the death of a significant person) or failure to come to terms with trauma (physical or sexual abuse). An Unresolved classification is assigned to a transcript if significant trauma or loss are described and there are indications of associated disorganization of discourse or disorientation of reasoning.

A fifth category, Cannot Classify (CC) has also been described (Hesse & Main, 2000). It represents a mixture of coexisting attachment strategies of dismissal and preoccupation within a single narrative. Previously such cases were rated in accordance with the closest single category (D, E, or F). It is likely that the CC classification will include individuals whose attachment representations are in some degree disorganized, and consequently they manifest multiple models in their narratives.

All the interviews were rated by two experienced and reliable raters who were blind to the sample and had no access to any information about the subjects other than gender. The current raters’ inter-rater reliabilities were consistent with those reported in the literature: 100% agreement for secure and insecure attachment classification, 85% on the major classification, and 70–80% on sub-classification categories (see Bakermans-Kranenburg & van IJzendoorn, 1993). Cohen’s kappa for the four-way classification was .8 between two raters.

2. The Reflective Function (RF) Scale (Fonagy, Steele, Steele, & Target, 1997). The RF Scale was devised by Fonagy et al. (Fonagy, Steele, et al., 1997) to use with AAI transcripts, to give a measure of the subject’s predisposition to infer mental states (feelings, thoughts,
beliefs, and intentions) in self and others in the course of attachment narratives. The RF Scale was originally based on Main’s (1991)\(^2\) metacognitive monitoring scale. The current coding system is described in a 50-page manual (Fonagy, Steele, et al., 1997) available from the second author. Coders are instructed to examine AAI transcripts with a view to establishing the clarity of the subject’s understanding of the nature of mental states as this emerges from descriptions of feelings and thoughts in the attachment figure and reactions to them. Answers to questions that specifically invite reflectiveness (e.g., “Why do you think your parents behaved as they did?”) are given extra weight. The lowest level of the scale is reserved for those who reject the invitation to be reflective (e.g., “You tell me why they behaved that way. You are the shrink.”)

The RF Scale is an 11-point scale, with manualized definitions for each anchor point. A low level of RF is assigned to narratives where generalizations or banal reflective statements dominate. Statements that appear to be accurate reflections of the thoughts and feelings of attachment figures or the subject in relation to them, but are partial (only occur occasionally in the interview), somewhat predictable, and adequate, rather than exceptional, indicate a moderate level of RF. Evidence for high RF is offered by consistent attention to psychological states of the protagonist in the narrative, making exceptionally frequent insightful statements on conscious and non-conscious motivations underpinning the subject’s own and others’ behaviour, particularly of attachment figures. The reliability of the RF scale after training is high \((r = 0.91)\).

3. **The Structured Clinical Interview for DSM–III–R (SCID I with psychotic screen)** (Spitzer, Williams, Gibbon, & First, 1990). This was used to establish current and past psychiatric disorder. This semi-structured interview provides detailed information on past and present psychopathology, for Axis I DSM–III–R diagnoses. It incorporates the Global Assessment of Functioning (GAF) Scale, which provides an overall measure of psychological, social, and occupational functioning, for Axis V of DSM–III–R. GAF provides a hypothetical continuum of mental health to illness on a 1–90 scale. To determine the reliability of the research psychiatrists on these instruments, Cohen’s kappa and Pearson’s \(r\) reliability coefficients were computed as appropriate. Values of were calculated for each Axis I diagnosis (SCID-I), yielding a median value of 0.85 (range 0.73–1.00).

4. **The Structured Clinical Interview for DSM–III–R Personality Disorders (SCID-II)** (Spitzer et al., 1990). This is another semi-structured interview that uses DSM–III–R operational criteria to diagnose personality disorders. The use of phenomenological assessments of PD are controversial (Westen, 1998), but the SCID-II is the most commonly used instrument for establishing whether patients meet DSM Axis II criteria. On Axis II (SCID-II), reliability of diagnoses was 0.61 for cluster A, 0.67 for cluster B, and 1.00 for cluster C. On the GAS an interclass correlation coefficient (ICC) of 0.78 was obtained for the total score, showing satisfactory inter-rater agreement.

5. **Forensic Psychiatric Clinical Interview.** A detailed clinical interview was carried out for each subject in the prison group in order to elicit information about the index and past offences, which were confirmed from Home Office files. There was generally a high level of agreement between prisoners’ reports and official records.

6. **National Adult Reading Test (NART)** (Nelson, 1982). This was used to estimate the

---

\(^2\) Is this referring to Main & Goldwyn, 1991? Yes
subject’s verbal IQ.

7. The Beck Depression Inventory (BDI) (Beck, 1993; Belsky, Rovine, & Taylor, 1984). This was used to assist in matching subjects on the severity of their depression. The BDI is a 21-item inventory that measures the subject’s symptoms and concerns about his or her current level of depression.

Results

1. Attachment Classifications

Table 2 shows the distribution of attachment classifications across the three samples. A chi-squared test revealed that the distributions were significantly different, with far fewer secure codings in the two psychiatric groups than in the medical controls ($\chi^2 = 23.3, df = 6, p < 0.001$). Contrasting the prisoners and the matched PD group, the differences remained statistically significant ($\chi^2 = 12.8, df = 3, p < 0.005$). More AAIIs were coded as dismissive in the prison sample than in the psychiatric group (36% compared to 0%), while the preoccupied classification was more common in the psychiatric controls than in the prisoners (50% compared to 14%).

A high proportion of subjects had CC as their primary classification, in both the prison and the PD control group, but this did not appear to differentiate the groups (32% as opposed to 27%). Examining the distribution of the sub-classification of the Ds classification revealed that in the prison sample the modal sub-classification was Ds1, with 63% of the dismissive subjects having this sub-classification. Ds1 represents the extreme end of the dismissive spectrum. In the PD control group the modal sub-classification was E2 (angrily preoccupied).

There were no differences in the prevalence of reported loss among the three groups. Abuse, severe enough to meet criteria, was reported only in the prisoners and the PD control sample (see Table 2). The prisoners reported significantly more abuse (82% versus 41%, Fisher Exact Probability Test $p < 0.05$). The distribution of type of abuse did not differ, with both groups being somewhat more likely to report physical rather than sexual abuse (Fisher Exact Probability Test, NS$^3$).

Qualitative examination of the transcripts revealed that not only did the prisoners report more abuse, but their reports frequently described extremely severe and highly bizarre punishments. One example was a prisoner who described how his father burnt his hands on an oven to punish him for stealing coins from a gas meter in his home. Another prisoner recalled being nailed to a cross, through his hands, as a young child. He had to be taken to hospital afterwards to have the nails removed. While in general these reports could not be verified, details such as the presence of scars on the subject’s hands in these cases lent support to the narratives.

Lack of resolution of mourning (LRM) scores above the cut point of 5 were given only in the two psychiatric groups, with none of the medical controls’ AAIIs meeting criteria for a U classification (see Table 2). Contrasting the two psychiatric groups, we found that somewhat fewer of the AAIIs of the prison group were coded Unresolved than in the PD control group (36% as opposed to 59%), but this difference did not reach statistical significance on Fisher’s Test. When we looked at lack of resolution to abuse separately from LRM for loss (see Table 2), we found that in the prisoner group lack of resolution to abuse was significantly less

---

$^3$ Should NS be ns, meaning non-significant? Yes
common (Fisher's Exact Test \( p < .05 \)).

2. AAI Scale Scores

In addition to examining differences in attachment classifications, we also explored differences between groups in terms of scores assigned to the AAI narratives on the experience and state of mind scales. Mean scores for the three groups are shown in Table 3. In the comparison between group means using one-way ANOVAs,\(^4\) highly significant differences were found on all the scales, with the exception of derogation and lack of recall. Post hoc tests with Bonferroni adjustment were performed separately on all scales between the prison and the two control groups. The differences between the prison and the medical groups were in the expected direction. All scales yielded significant F values. The comparison of the prisoner and personality-disordered groups on the experience scales yielded significant differences only for “pressure to achieve” and “neglect” ratings, with the prison group rated lower on the former and higher on the latter. On the state of mind scales, AAIs for the prison group were rated as less angry, more idealizing, and less coherent than the PD controls.

<Table 3 here>

3. Reflective Function Scale Scores

The mean reflective function ratings for the AAI transcripts of the three groups are also shown in Table 3. The one-way ANOVA was highly significant, with both psychiatric groups scoring relatively low, while the mean for the medical control group was close to the expected value for a normal sample. On post hoc comparison, the prison group emerged as significantly lower in reflective function than the PD psychiatric controls (\( p < .005 \)).

We further compared the clinical groups in terms of the number of subjects with RF in the “deficit” range, defined as a rating below 3. Of the prison group, 73% had low RF by this definition, compared with 32% of the PD psychiatric controls (Fisher Exact Test \( p < 0.01 \)).

Finally, while the groups were relatively well matched for Axis II disorders, anxiety disorder was more common in the psychiatric control group. We wanted to ascertain that differences in the distribution of attachment classifications and RF could not be accounted for in terms of the under-representation of anxiety diagnoses in the prison sample. A three-way contingency table was created with high versus low anxiety, prison versus PD psychiatric group, and attachment classification respectively as factors in the analysis of attachment classification. The hierarchical log linear model yielded no main effect interaction with anxiety for either dismissing or preoccupied AAI classifications. The analysis was repeated with high–low RF level as the third variable with similar results.

The prison sample was subdivided according to their index offences, and into a violent and non-violent subgroup. Violence against attachment figures dominated the offences in the violent group. The victims of 4 of the most violent offences (murder, wounding with intent, and G.B.H.) were in a current attachment relationship with the perpetrator. The victims of the 3 sex offences were also known to the offenders. In contrast, victims of the non-violent offences were not known to the perpetrator.

Fisher’s Exact Test was used to examine whether individuals with violent offences were more likely to be rated below 3 on RF. Of individuals with violent offences, 93% had low RF scores, in contrast to 29% of those with non-violent offences (\( p < .004 \)). The mean RF ratings obtained by the violent group was also significantly lower on a \( t \)-test (Mean\(_{VIOL} = 1.40\);

\(^4\) Perhaps this initialism should be spelled out on first use. *Analysis of variance*
Mean_{NON-VIOL} = 2.82; t = 2.38; DF = 21; p < .02.

Discussion

Adult attachment interviews with this group of psychiatrically disordered prisoners have shown that this group are distinct in their representations of attachment relationships. Compared to both personality-disordered and a normal control group, they were significantly more likely to be insecure in their attachment classifications. The prison and PD group were more likely to be insecure than a control group without major psychiatric problems. In terms of the form of insecurity, prisoners were more likely to be dismissive than a control group with the same PD diagnosis. An unexpectedly high proportion of AAIs were classified as Ds1, which represents the extreme end of the dismissive category, characterized by idealization and a failure to acknowledge attachment-related difficulties. The greater number of dismissive cases in the prison group is consistent with Bowlby’s view that offending behaviour is associated with distinct and abnormal patterns of attachment. The observation that offenders were more likely to be coded as dismissive (particularly the extreme Ds1 category) supports Bowlby’s hypothesis that offending is a consequence of the disruption of early attachment experiences, in particular neglect and severe physical abuse (Main, Kaplan, & Cassidy, 1985). Nevertheless, it should be remembered that only 36% of AAIs were coded as dismissing. However, an additional 32% were classified as CC which carries significant dismissing features.

The prisoners reported more experiences of abuse than the PD psychiatric controls. Type of abuse did not differentiate the groups, although there appeared to be somewhat more physical abuse in the prisoners, sometimes quite severe and bizarre. The findings of abuse are, however, based on self-report in the AAI rather than standardized interviews designed for the purpose of evaluating childhood maltreatment (Bifulco, Brown, & Harris, 1994). However, the prisoners consistently reported experiences that reflected less concern with them as children (higher rates of neglect and less pressure to achieve). Despite this lack of concern and more experiences of abuse, their state of mind related to attachment appears as less angry, more idealizing, and less coherent than that of a matched control group of individuals with very similar psychiatric and personality-disordered diagnoses, but without an offending history. A psychoanalytic understanding may conceptualize the prisoners’ tendency to be more dismissive, less angry at and more likely to idealize one parent, despite their experiences of maltreatment, in terms of splitting (Kernberg, 1992). They split their internal parental objects into good and bad, and use processes of denial, idealization, and identification with the aggressor to protect their feelings about their good internal object and themselves.

Surprisingly, although the prisoners reported histories of severe and sometimes bizarre physical abuse, on the AAI they rarely met criteria for lack of resolution of these experiences. Disorganized discourse or disorientated reasoning in relation to the trauma is required as evidence for the unresolved classification in the Main and Goldwyn system (Main & Goldwyn, 1998). We propose that the prisoners instead disavow (Freud, 1927/1961) their attachment experiences, so that they do not express them in the form recognized by this classification system. The disavowal limits their capacity for mental representation of such experiences, including those of trauma, that are instead experienced concretely and even physically,
making them liable to act in bodily ways and offend.

In contrast, the PD psychiatric group were rated as more unresolved in relation to their abuse. They were also more likely to be preoccupied in their attachment patterns (a high proportion were rated E2 sub-category of preoccupied status, characterized by anger towards caregivers). This finding supports earlier studies that have shown a predominance of preoccupied attachment patterns in BPD patients (Fonagy et al., 1996; Patrick, Hobson, Castle, Howard, & Maughan, 1994). They had more anxiety disorders, which have been found to be associated with unresolved states (Fonagy et al., 1996). We propose that the personality-disordered patients, who are more unresolved to their trauma, more anxious and angry about it, are more likely to be recognized as requiring treatment. We hypothesize that the prisoners, with their dismissive strategy, apparent lack of unresolved states, low anxiety, and offending behaviour may be more liable to be judged negatively by society, requiring punishment rather than treatment. We therefore suggest that selection biases could be operating, so that patients who express more unresolved states to their trauma and have a greater capacity to reflect are more likely to be offered treatment.

Finally, we found that prisoners were significantly lower in reflective function ratings than the PD control group, whose RF was in turn lower than that of the normal controls. Those prisoners whose offences were violent in nature were rated lowest on this scale. Their low reflective function confirms the hypothesis that the prisoners have a limited capacity to think or “mentalize” about their own or others’ mental states, in terms of feelings, thoughts, beliefs, or motivation. We suggest that severe early trauma in the context of attachment experiences leads to a developmental line of psychopathology, characterized by a disavowal of attachment experiences and capacity to think about them, resulting in a deficit of RF and meta-cognitive thinking. The subject has then a predisposition to experience mental states in physical and bodily ways, making him or her liable to offend, especially through violent acts.

The study has major limitations. These include the small sample size, the cross-sectional case control design, and the limited assessment of the medically ill control group. This is not an epidemiological sample, and generalizations are therefore limited. It would be important to explore whether violent offenders who appear less reflective than non-violent offenders were also likely to be more dismissing and less preoccupied and angry. However, the sample size of violent offenders was far too small to permit a legitimate contrast on the categorical variables related to attachment. The study has not explored the predictors of RF, which may have generated alternative accounts. The sample was too small and the available measures too sparse to put alternative accounts to substantive test. The therapeutic implications of our observations at this stage can only be speculative. Even if low RF turns out to be on the causal path to violence, this does not necessarily imply that addressing this issue therapeutically would lead to a reduction in violent acts.

The common path to violence is a momentary inhibition of the capacity for mentalization. This requires one of three conditions: (1) a particular biology in which intentional states are not normally responded to by the individual (“I cannot recognize”); (2) a particular personal history in which the person cannot recognize intentional states because his or her intentional states were not normally responded to (“I am not recognized”); (3) a particular social environment in which the individual feels merged with other subjectivities, and the biological need to see self and others as intentional is temporarily removed, as might happen in a large group or as part of military training (“I cannot be recognized”). Each mode of violence
demands a different degree of inhibition of mentalization, depending on factors such as (1) felt anonymity, (2) the physical proximity of the victim, (3) the time it takes to carry out the act, and (4) the amount of eye contact that the violent act entails, since it is through the eyes that intentional states are normally read (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). Modes of violence requiring less inhibition of mentalization are likely to be practised by a wider range of people in a wider range of contexts and are therefore more dangerous (thus using one’s hands takes longer than using a knife, which involves more proximity than the use of a gun, which in turn requires more inhibition of mentalization than the use of a bomb). This formulation considers low RF to be a “trait” variable predisposing an individual to the complete loss of mentalizing in the context of high levels of arousal.

There are two groups of individuals for whom this evolutionary design proved ineffective. The first group are likely to be there because of genetic predisposition rather than social experience (Moosajee, 2003; Sluyter, Arseneault, Moffit, Veenema, de Boer, & Koolhaas, 2003). Thus individuals constitutionally poor at recognizing mental states in others through facial expressions or vocal tones may not fully acquire mentalization and thus inhibit their natural violence (Blair, 2001; Blair & Cipolotti, 2000; Blair, Morris, Frith, Perrett, & Dolan, 1999). In line with the terrible threat such individuals represent, we dismiss them as psychopaths, a term intended to create maximal distance between them and us.

Individuals in the second group may not acquire the capacity to interpret minds simply because they never had the opportunity to learn about mental states in the context of appropriate attachment relationships. Alternatively, their attachment experiences may have been cruelly or consistently disrupted. For others, the emerging capacity for mentalization has been destroyed by an attachment figure, whose thoughts and feelings about the child provoked sufficient anxiety for the child to want to avoid thinking about these thoughts and feelings. We have claimed along with others that the capacity for mentalization is linked to attachment (Fonagy, Gergely, Jurist, & Target, 2002; Meins, Ferryhough, Fradley, & Tuckey, 2001). We learn about minds, our own and those of others, through experiencing our internal states being understood by another mind. At least three types of dysfunctions of attachment may lead to violence: (1) attachment experiences may have been consistently disrupted by combination of social circumstance and parental failure; (2) attachment problems associated with child’s temperament (e.g., fearlessness leading the child not to seek out attachment figure, in turn leading to a failure to acquire the capacity robustly to mentalize); (3) later attachment trauma when a nascent capacity for mentalization has been destroyed by a powerful figure, who created sufficient anxiety about his or her thoughts and feelings towards the child for the child to wish to avoid thinking about the subjective experience of others (see Fonagy, Target, Steele, & Steele, 1997; Fonagy, Target, Steele, Steele, Leigh, Levinson, et al., 1997). It is important to retain an awareness of the possibility that violence may be rooted in the disorganization of the attachment system. A child may manifest an apparent callousness that is actually rooted in anxiety about attachment relationships. Perhaps this is also part of an evolutionarily adaptive scenario, because a harsh early childhood could signal greater future need for interpersonal violence (see Belsky, 1999). Our findings are consistent with psychoanalytic formulations offered by clinicians with forensic interests, for instance Meloy and Sohn, who also link violence to the projection of an intolerable aspect of the self,
including unbearable affect and its projective identification in the victim, who is attacked to destroy this aspect (Meloy, 1992, 2000; Sohn, 1995), and are also consistent with Sohn's (1995) clinical description of patients with a history of trauma and loss, who are unable to tolerate the experience of depression and humiliation. In such individuals a further rejection may trigger a violent act, because they lack the capacity for symbolization.

The group whose aggression is high in early childhood, and continues into adolescence and early adulthood, we argue are likely to have had attachment experiences that failed to establish a sense of the other as a psychological entity. We know from other longitudinal work (Rutter, 2000) that environmental influences that divert the child from paths of violence and behavioural disturbance often imply the establishment of strong attachment relationships with relatively healthy individuals. In such relationships the adolescent can acquire implicit knowledge of minds. The provision of psychotherapy for disturbed adolescents and adult offenders may assist in this developmental change. To reduce the risk of violence to us, we need to ensure that social institutions (families, nurseries, schools, detention centres, prisons, and secure hospitals) are designed to enrich representations of mental states in self and others. For example, teachers should help the class to reflect on incidents of bullying rather than adopt power assertive strategies of exclusion (Twemlow et al., 2001; Twemlow, Fonagy, Sacco, Otoole, & Vernberg, 2002). Prisons that adopt power-assertive strategies as part of a social defence system, activating organizational defence mechanisms of splitting and projection, too often become a re-enactment of the offenders’ early attachment experiences. The splitting and projection of good and bad, of power into the authorities, of helplessness into the inmates, idealization of rigid responses and the denigration of thinking promote further brutalization, humiliation, and aggression. To reduce the risks of trans-generational transmission of this type of psychopathology and the escalation violence, prisons and centres for adolescents could be developed along psychotherapeutic lines, incorporating a “culture of enquiry” (Main, 1983) to reflect upon the institutional dynamics and individuals’ problems and psychotherapy to help individuals develop a capacity to reflect on their own and others’ minds.

### Table 1. Demographic and Clinical Characteristics of the Samples

<table>
<thead>
<tr>
<th></th>
<th>Prison group</th>
<th>Personality disorder controls</th>
<th>Normal controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects (n)</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Demographic variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (% male)</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Mean age (s.d.)</td>
<td>28.9 (8.3)</td>
<td>30.3 (5.8)</td>
<td>31.5 (4.9)</td>
</tr>
<tr>
<td>Mean estimated IQ (s.d.)</td>
<td>110.9 (10.5)</td>
<td>113.4 (8.9)</td>
<td>112.8 (8.4)</td>
</tr>
<tr>
<td>Social class (% SC IV &amp; V)</td>
<td>95%</td>
<td>70%</td>
<td>75%</td>
</tr>
<tr>
<td>% non-white</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
</tr>
<tr>
<td>Diagnostic data (% meeting DSM criteria)</td>
<td>77%</td>
<td>96%</td>
<td>NA</td>
</tr>
<tr>
<td>Anxiety disorder*</td>
<td>5%</td>
<td>32%</td>
<td>NA</td>
</tr>
</tbody>
</table>

6 What does the initialism s.d. stand for? **Standard deviation**
Substance abuse 77% 60% NA
Eating disorder 0% 9% NA
Cluster B, PD 80% 69% NA
Other PDs 59% 64% NA
Mean No. Axis I diagnoses 3.0 (2.3) 3.1 (1.2) NA
Mean No. Axis II diagnoses, (s.d.) 3 (2.3) 2.5 (0.9) NA
Mean GAF score (s.d.)** 48.3 (7.3) 32.3 (10.1) NA
Mean BDI (s.d.) 22.31 (14.1) 24.1 (10.0) NA

*significant difference p < 0.05. **significant difference p < 0.011

Table 2. Distribution of AAI Classifications across the Three Groups

<table>
<thead>
<tr>
<th></th>
<th>Prison group n (%)</th>
<th>Personality disorder controls n (%)</th>
<th>Medical controls n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four-way classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secure</td>
<td>4 (18.2%)</td>
<td>5 (22.7%)</td>
<td>12 (54.5%)</td>
</tr>
<tr>
<td>Preoccupied</td>
<td>3 (13.6%)</td>
<td>11 (50%)</td>
<td>3 (13.6%)</td>
</tr>
<tr>
<td>Dismissing</td>
<td>8 (36.4%)</td>
<td>0</td>
<td>5 (22.7%)</td>
</tr>
<tr>
<td>CC</td>
<td>7 (31.8%)</td>
<td>6 (27.3%)</td>
<td>2 (9.1%)</td>
</tr>
<tr>
<td>Abuse history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>14 (63.6%)</td>
<td>8 (36.4%)</td>
<td>0</td>
</tr>
<tr>
<td>Sexual</td>
<td>4 (18.2%)</td>
<td>1 (4.5%)</td>
<td>0</td>
</tr>
<tr>
<td>Neither</td>
<td>4 (18.2%)</td>
<td>13 (59.1%)</td>
<td>22 (100%)</td>
</tr>
<tr>
<td>Unresolved with respect to loss or trauma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolved</td>
<td>14 (63.6%)</td>
<td>9 (40.9%)</td>
<td>22 (100%)</td>
</tr>
<tr>
<td>Unresolved</td>
<td>8 (36.4%)</td>
<td>13 (59%)</td>
<td>0</td>
</tr>
<tr>
<td>Unresolved with respect to physical or sexual abuse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolved</td>
<td>12 (66.7%)</td>
<td>2 (22.2%)</td>
<td>NA</td>
</tr>
<tr>
<td>Unresolved</td>
<td>6 (33.3%)</td>
<td>7 (77.8%)</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 3. Mean and Standard AAI Scale Scores of Prisoners and PD and Medical Controls

<table>
<thead>
<tr>
<th>GROUPS</th>
<th>1. Prison group (n = 22) mean(SD)</th>
<th>2. PD control (n = 22) mean(SD)</th>
<th>3. Medical control (n = 22) mean(SD)</th>
<th>ANOVA (F)</th>
<th>Bonferroni adjustment p values for post hoc controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience Scale</td>
<td>Group 1 vs. Group 3</td>
<td>Group 1 vs. Group 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loving</td>
<td>2.66 (0.8)</td>
<td>2.30 (1.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rejection</td>
<td>5.84 (1.6)</td>
<td>4.49 (1.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neglect</td>
<td>6.94 (1.3)</td>
<td>6.21 (1.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pressure</td>
<td>1.30 (0.6)</td>
<td>2.29 (0.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State of mind Scale</th>
<th>Group 1 vs. Group 3</th>
<th>Group 1 vs. Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>2.64 (1.6)</td>
<td>2.30 (1.4)</td>
</tr>
<tr>
<td>Idealisation</td>
<td>4.15 (1.9)</td>
<td>4.49 (1.3)</td>
</tr>
<tr>
<td>Derogation</td>
<td>2.13 (1.3)</td>
<td>2.29 (0.9)</td>
</tr>
<tr>
<td>Passivity</td>
<td>4.02 (2.4)</td>
<td>2.29 (0.9)</td>
</tr>
<tr>
<td>Coherence</td>
<td>3.16 (1.7)</td>
<td>2.29 (0.9)</td>
</tr>
<tr>
<td>Fear of loss</td>
<td>1.59 (1.0)</td>
<td>2.29 (0.9)</td>
</tr>
<tr>
<td>Recall</td>
<td>4.19 (2.5)</td>
<td>2.29 (0.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reflective function</th>
<th>Group 1 vs. Group 3</th>
<th>Group 1 vs. Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.11 (1.4)</td>
<td>2.29 (0.9)</td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05. ** p < 0.01. *** p < 0.001

REFERENCES


---

7 Is this *Journal of Neurology, Neurosurgery, and Psychiatry*? Yes
8 Should this be a smaller number? See previous entry. Should be 122 (Pt 5)
9 Should this be *Parent–Child*? Yes
10 Is this *Journal of the Personality and Social Psychology*? Yes
11 Is this *Criminal Behaviour and Mental Health*? Yes


12 What does this initialism stand for? Her Majesty’s Stationery Office


\[\text{Is this } \textit{Criminal Behaviour and Mental Health}? \ \text{Yes}\]

\[\text{Is this } \textit{Criminal Behaviour and Mental Health}? \ \text{Yes}\]

\[\text{In other versions, the word is system. Should this be singular as well? } \text{Yes}\]


17 Infant’s? Yes

18 City? Northvale

19 Is this *Developmental Psychopathology*? Yes

20 Is this *Journal and Review of Social Medicine*? *Journal of the Royal Society of Medicine*
school age conduct problems. Unpublished manuscript.

Alice Levinson
Epping Forest Psychotherapy Service
Peter Fonagy
Sub-Department of Clinical Health Psychology
University College London
Gower Street
London WC1E 6BT

21 Is this Behaviour and Genetics? Behavior Genetics
22 Is this Journal of Child Psychology and Psychiatry? Yes