Facial responses to facial expressions of emotion in dangerous and severe personality disorder

Keith William Whittle

Submitted in accordance with the requirements for the degree of
Doctor of Clinical Psychology (DClinPsychol)

The University of Leeds
Academic Unit of Psychiatry and Behavioural Sciences
School of Medicine

June 2006

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

ACKNOWLEDGEMENTS

It would not have been possible to complete this research without the help and support of a number of people. Much of my gratitude goes to the participants, both at HMP Whitemoor and at the University. I am indebted to Dr Naomi Murphy, for helping to organize access to prisoners, for letting me know which people I needed to speak to and when, and for facilitating all aspects of my visits to the prison. Thanks to all the staff at HMP Whitemoor, from the officers and psychology staff on D Wing and A Wing to the people on the gate, for their friendliness and professionalism. A special thanks to Sophie Strong, for ensuring that many of my visits to the prison ran smoothly. To Dr Mitch Waterman, I am extremely grateful – for being 100% behind the research the whole time, and for providing me with intellectual, moral, *and* emotional support throughout the whole process. Katie and Louise provided me with much needed technical guidance and the occasional dinner. Thanks to my family for the well-timed phone calls. And finally, thank you Jenny, for managing to live with me for the past twelve months.

ABSTRACT

When people view facial expressions of emotion they tend to mimic or mirror the expression. According to Leventhal's perceptual-motor theory of emotion, this response occurs when innate central motor programmes are activated. There is evidence that individuals with psychopathic traits have impaired autonomic responses to and recognition of emotional expressions. There is further evidence that individuals who lack empathy show reduced mirroring to facial expressions of emotion. This research piloted a method to investigate the facial responses of a group of eleven prisoners on the Dangerous and Severe Personality Disorder (DSPD) Programme at HMP Whitemoor. Participants in this group had elevated Psychopathy Checklist – Revised (PCL-R) scores (*Mean* = 28.4, *SD* = 3.6). Two control groups were recruited: prisoners who were not on the DSPD Programme (N = 9), and a group of university employees and students (N = 10). Participants were filmed as they completed a facial affect recognition task using dynamic, spontaneous facial expressions of emotion as stimuli. The presence or absence of mirroring was determined by two independent raters. Because interrater reliability was low (*Mean Cohen's* κ = 0.28, *SD* = 0.15), ratings were analysed separately. Non-parametric tests were used to investigate differences in group means for all analyses except for recognition, as only these data met parametric assumptions. There were no group differences in mirroring at the 5% level. However, group differences between university controls and prison controls approached significance, with more mirroring of *happiness* (both raters) and *disgust* (one rater only) in the university controls. Differences between the DSPD group and university controls on these measures were also (nonsignificantly) in the predicted direction. Supplementary analyses found higher recognition of *anger* in the normal controls than in the prison controls. There was no association between mirroring and recognition, and there were no group differences in emotional sensitivity based on strength ratings for the stimuli; however, university participants selected more emotional classes per trial than either of the two prison groups. It is proposed that in order to measure possible deficits in both mirroring and recognition, it may be necessary to use a combination of extreme and moderate facial stimuli, balanced across emotional classes. The potential application of this method to the assessment of mirroring in DSPD, and also to the evaluation of treatment, is discussed.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AU	Action unit
APD	Antisocial personality disorder
BPD	Borderline personality disorder
DoH	Department of Health
DSM	Diagnostic and Statistical Manual of Mental Disorders
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition – Text Revision
DSPD	Dangerous and severe personality disorder
EMG	Electromyography
FACS	Facial Action Coding System
HMP	Her Majesty's Prison
HO	Home Office
PCL-R	Psychopathy Checklist – Revised
PCL-SV	Psychopathy Checklist – Short Version
PD	Personality disorder

INTRODUCTION

This study brings together research on *psychopathy*, *antisocial personality disorder* and *Dangerous and Severe Personality Disorder (DSPD)*, research looking at facial responses to facial expressions of emotion, and the facial affect recognition task. The research offers contributions that are consistent with two of the three priorities that have been set by the Home Office for DSPD research: *improve knowledge around the causes of Personality Disorder (PD)*, and *evaluate specific DSPD assessment and treatment programmes*. A practical method for assessing emotional recognition and facial responses is piloted. It is hoped that from this research mechanisms may be inferred that contribute to our understanding of DSPD. From a clinical perspective, this research may suggest areas of deficit in individuals who meet the criteria for DSPD, while the method could be developed to evaluate the outcomes of treatments that target these deficits.

LITERATURE REVIEW

Psychopathy

According to Cleckley, author of arguably the most influential book on the subject of psychopathy and the first person to describe the label in clinical terms, psychopaths are rationally and cognitively intact individuals, who are unable to appreciate the emotional significance of human behaviour (*The Mask of Sanity*; Cleckley, 1976). Psychopaths are characterized by Cleckley as callous, possessing a diminished capacity for remorse, superficial charm, a lack of empathy, proneness to boredom, and impulsivity.

In the *Diagnostic and Statistical Manual: 4th Edition Text Revision* (DSM-IV-TR; American Psychiatric Association, 2000), *antisocial personality disorder* (APD) is firmly related to, but not synonymous with, psychopathy. Among adult prisoners, the vast majority of individuals who are psychopathic according to the PCL-R meet the criteria for APD, while only approximately 25% who meet the criteria for APD are also designated as psychopaths (Hare, Hart, and Harpur, 1991). Psychopathy is in fact defined as a score of 30

or above on the *Hare Psychopathy Checklist – Revised* (PCL-R; Hare, 1991), a clinician-scored rating scale developed by Robert Hare and based in large part on Cleckley's original conceptualisation.

In the PCL-R manual, Hare (1991) proposed that a score of 30 or above (out of 40) should be used to classify psychopaths, corresponding to a score around the 75th percentile for offenders. However, researchers have found that lower cutoffs can provide a more optimal division of groups of individuals into psychopaths and nonpsychopaths. For example, Harris, Rice, and Quinsey (1994) reported that a PCL-R cutoff of 19-20 maximised the hit-rate (minimising false positives and false negatives) in a sample of 653 serious offenders, while applying a cutoff of 25 yielded a "nearly pure" subsample (Harris et al., 1994; p395). In an earlier study, Harris, Rice, and Cormier (1991) indicated that in a group of 169 adult male offenders with mental health problems, a cutoff of 30 did not improve the prediction of reoffending over a cutoff of 25. Cooke and Michie (1997) used item response theory to analyse PCL-R data from a sample of 2,067 North American prisoners and forensic patients, and concluded that discrimination was maximized by this reduced cutoff. Moreover, in a study comparing North American and Scottish forensic samples, Cooke and Michie (1999) found that in order to establish metric equivalence between the two samples, it was necessary to reduce the cutoff for the Scottish sample to 25.

Initial factor analyses of behaviours rated on the PCL-R revealed two independent factors: an emotion dysfunction factor defined largely by emotional shallowness and lack of guilt, and an antisocial behaviour factor defined largely by instrumental aggression and criminal behaviours (Hare et al., 1991). APD is more closely related to the latter, behavioural factor. The diagnosis largely ignores the affective and interpersonal characteristics of Factor 1, helping in part to explain the diagnostic asymmetry between the two labels. The two psychopathy factors have been emphasised in much of the subsequent literature; however, in a substantial review and re-analysis of the data Cooke and Michie (2001) concluded that three factors give a better description of the data: Arrogant and Deceitful Interpersonal Style, Deficient Affective Experience, and Impulsive and Irresponsible Behavioural Style. This revision received some criticism for its exclusion of items relating to antisocial

behaviour, and in the recently published manual to the 2nd Edition of the PCL-R, Hare appears to acknowledge both Cooke and Michie's refinement in addition to these concerns by proposing a 4-factor model (Hare, 2003). The critical difference with Hare's model is the inclusion of an antisocial factor, in addition to the interpersonal, affective, and impulsive/irresponsible factors.

The PCL-R has been found to be both reliable and valid in a number of studies across several populations (e.g., Hare, Harpur, Hakstian, Forth, Hart, and Newman, 1990, analysed data from five prison samples; Vitale and Newman, 2001, reviewed the use of the PCL-R with female institutionalized and non-institutionalised samples; Moltó, Poy, and Torrubia, 2000, looked at a Spanish prison sample; for reviews, see Salekin, Rogers, and Sewell, 1996; Hare, 2003). The PCL-R has been shown to have good predictive validity for a variety of antisocial behaviors, including criminal violence (Cooke and Michie, 1997), violent recidivism following release from prison or hospital (Harris, Rice, and Quinsey, 1993), and poor treatment response to correctional treatment programs (Ogloff, Wong, and Greenwood, 1990).

One of the main criticisms of the PCL-R is its lack of applicability to non-forensic populations, primarily because of the inclusion of behavioural criteria that apply exclusively to prisoners – specifically "revocation of conditional release" and "criminal versatility". The instrument has also been criticized for its use of the term *psychopath*, a label some have argued carries emotional baggage (Gendreau, Goggin, and Smith, 2002). On a more practical note, Gendreau et al. argue that it can be difficult and cumbersome to use, requiring up to two hours to administer and graduate level of qualification. With respect to the former point, it could be argued that the negative connotations of the term *psychopath* stem more from the behaviours and affective components that underpin the label, rather than because of the term itself. According to this reasoning, changing the label would simply lead, in time, to a new negatively-charged term for these individuals. On the latter point, there may be some tacit acknowledgement (in the face of explicit denials, for example, Hemphill and Hare, 2004) from Hare that the time required for administration is considerable, with the development of the shorter *PCL – Short Version* (PCL-SV; Hart,

Cox, and Hare, 1995). Nonetheless, contentions and limitations notwithstanding, the PCL-R has been used in more psychopathy research than any other measure and – particularly when it is used with the forensic populations on which it has been most thoroughly validated – is generally recognized as the most valid and useful instrument to assess psychopathy (Fulero, 1995; Stone, 1995).

Personality disorder

In addition to antisocial personality disorder, which is associated with psychopathy, there are nine other personality disorder diagnoses included in Axis II of DSM-IV-TR. These are arranged into three clusters based on theoretically similar underlying pathology. Cluster A personality disorders (paranoid, schizoid, and schizotypal) are characterized by odd or eccentric behaviours, Cluster B (antisocial, borderline, histrionic, and narcissistic) by dramatic or erratic behaviours, and Cluster C (avoidant, dependent, and obsessive-compulsive) by anxious or inhibited behaviours. There are also two additional personality disorders in an appendix: depressive personality disorder and passive-aggressive personality disorder. Pending further empirical support, these latter two disorders may appear in the main body of DSM-V. Finally, there is a diagnosis of personality disorder not otherwise specified, for cases which come close to meeting the requirements of one or more personality disorders, but do not fulfill the sufficient number of criteria in any one disorder. Diagnosis, as with Axis I disorders, is categorical and primarily comprises behaviours. In contrast with the major mental disorders of Axis I, however, there is an age restriction insofar as onset must be no later than early adulthood.

One of the features purported to distinguish the personality disorders from mental illness is that the former are held to be present and relatively stable throughout adulthood. However, it has become increasingly apparent that the same can be said for a proportion of the cases of mental illnesses, e.g., schizophrenia (Shea and Yen, 2003). Conversely, there is evidence that personality traits can in fact change over time, to the extent that a person may meet the criteria for a personality disorder at one stage in life, but not at a later stage, without a

therapeutic intervention (Shea, Stout, Gunderson, Morey, Grilo, McGlashen, Skodol, Dolan-Sewell, Dick, Zanarini, and Keller, 2002). The distinction between Axis I and Axis II disorders is confused further by the fact that the presence of an Axis I condition, such as depression, can impact on whether a personality disorder is diagnosed (Shea and Yen, 2003).

Perhaps the greatest conceptual challenge facing personality disorders – as they are currently conceived – is the fact that the presence of features from multiple personality disorder diagnoses are the rule for individual patients, rather than the exception (Oldham, Skodol, Kellman, Hyler, Rosnick, and Davies, 1992). Alternatively, people may exhibit features from a number of different diagnoses, while not fulfilling the required number of criteria for any one. One of the aims during the development of the classification system was to make the diagnoses as distinguishable as possible; however, in creating the prototypical syndromes of the current DSM, the (misleading) impression is created that pure types should regularly be seen in clinical practice.

Dangerous and Severe Personality Disorder

The term Dangerous and Severe Personality Disorder (DSPD) was first used in a British Department of Health (DoH) and Home Office (HO) paper, *Managing Dangerous People with Severe Personality Disorder* (DoH, HO, 1999), which made proposals to detain and treat a small minority of offenders. It is not at present a clinical or legal term, but is described as a "working definition to describe the very small group of people with a severe personality disorder who, because of their disorder, also pose a significant risk of serious harm to others" (Warren, Preedy-Fayers, McGauley, Pickering, Norton, Geddes, and Dolan, 2003, p4).

A year after the release of the DoH/HO paper, the Government's proposals to manage dangerous people with severe personality disorder were to include a research programme to build a sound evidence base from which to develop DSPD services. A large review of the

literature on treatment of psychopathy and personality disorders had concluded that the evidence for treatability was limited to a small number of studies, themselves limited by poor methodology (Dolan and Coid, 1993). In a subsequent review, taking into account the period from 1993 to 2001 in addition to the data covered by Dolan and Coid, similar conclusions were reached and the quality of evidence for the treatment of severe and/or dangerous personality disorders in particular was found to be poor (Warren et al., 2003).

In an effort to address the empirical gaps in our knowledge an expert group, set up to advise the DSPD programme, set out the following three priorities for DSPD research:

1. Improve knowledge around the causes of Personality Disorder (PD)
2. Explore the relationship between PD and serious offending behaviour
3. Evaluate specific DSPD assessment and treatment programmes

Four new specialist services have now been set up as pilots around England to treat offenders who have been assessed as having a dangerous and severe personality disorder: HMP Whitemoor, HMP Frankland, Broadmoor Hospital, and Rampton Hospital. HMP Whitemoor was the first, and is currently developing and constantly evaluating a psychological treatment model based on a four-stage treatment model incorporating the following sequence of phases: *problem recognition*; *exploration*; *acquisition of alternative behaviours*; and *consolidation and generalization* (Livesley, 2001). An additional stage, *engagement*, has been added to the model at HMP Whitemoor. This is intended to highlight the fact that more emotionally damaged individuals can find it much more difficult to engage in the therapeutic alliance, with particular emphasis on moving beyond a superficial commitment required for meeting targets set by a parole board, towards gaining an understanding of what it means to engage emotionally or to have a meaningful, internally driven desire for change.

Empathy

The word *empathy*, a construct believed to be fundamentally lacking in the psychopath (Hare, 1991) can be traced back to the German word *Einfühlung*, which was put forward by Robert Vischer (1873; in Listowel, 1934) to mean the spontaneous projection of real psychic feeling into the people and things they perceive. Lipps (1903; in Wispe, 1987) developed the theory, suggesting that people knew and responded to each other through *Einfühlung*, which was preceded by projection and imitation, and that as imitation of affect increases *Einfühlung* increases. The English word *empathy* was coined in 1909 by Titchener and defined as a "process of humanising objects, of reading or feeling ourselves into them" (Titchener, 1924). According to Leventhal's (1984) hierarchical perceptual-motor theory of emotion, this imitation would be biologically prepared in the form of innate central motor programs (the first level of the model). Interactive experiences between an individual and other people throughout development would feed into the second level, that of schematic motor codes, which would serve to enhance a person's imitative responses (see Fig. 1).

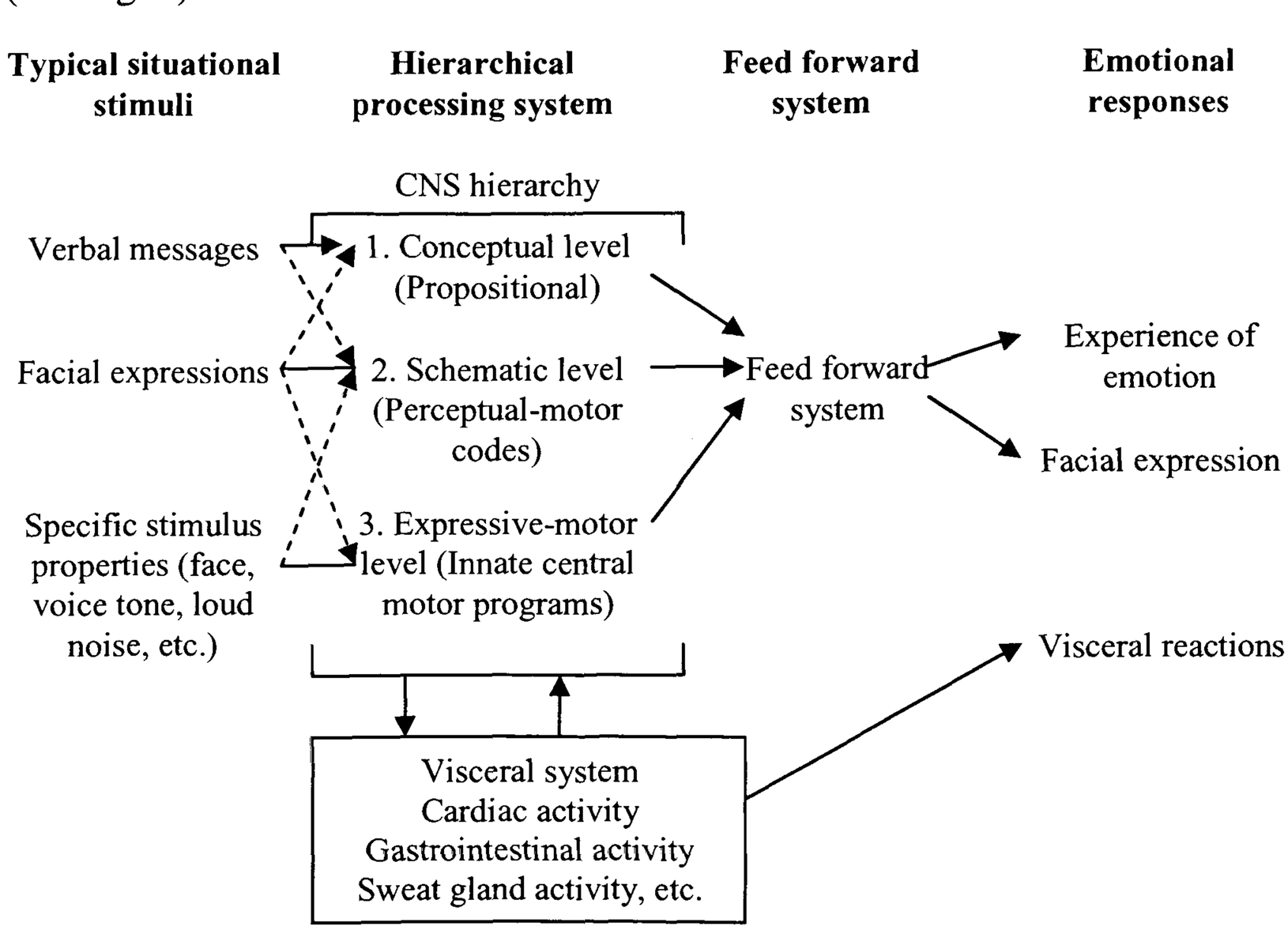

Fig. 1: Model of the hierarchical processing system for the construction of emotional reactions (based on Leventhal, 1984).

Leventhal stresses that while peripheral or motor activity, including facial-motor, is not necessary for emotional experience and behaviour, such activity does feed back into and is integrated with emotional processing with the usual effect of intensifying emotional experience.

Measuring empathy

In their review of empathy research Duan and Hill (1996) list several measures, including those that focus on the purely cognitive (e.g., Hogan, 1969; Truax and Carkhuff, 1967), those that conceptualise empathy as affective (e.g., Eisenberg, Fabes, Bustamante, and Mathy, 1987; Feschbach and Roe, 1968; Mehrabian and Epstein, 1972) and a measure that takes both cognitive and affective aspects into account (Davis, 1980). Duan and Hill question the validity of the available measures. They observe that self-report measures are limited by "human perception errors" (p. 264), whereby a person may believe that they understand or are understood by another, but this may in fact not be reciprocated. Observer ratings, on the other hand, focus on the outward expression of what is taken to be an inner experience, and such methods may be confounded by the clinician's communication skills.

There is empirical support for a lack of empathy in psychopathy using standard empathy measures, e.g., Zagon and Jackson (1994) found in a group of 149 students that, overall, those scoring higher on a measure of psychopathy scored lower on a measure of empathy. In a study of conduct-disordered adolescents Cohen and Strayer (1996) investigated the relationship between empathy and conduct disorder, a disorder defined by DSM-IV-TR as a necessary but not sufficient precursor of antisocial personality disorder. They found that empathy was lower among conduct-disordered than comparison adolescents and was related inversely to antisocial and aggressive attitudes. And finally, in a review of studies looking at the association between empathy and aggressive or antisocial behaviour – characteristics of Hare's Factor 2 – Miller and Eisenberg (1988) found them to be negatively related.

Facial responses to facial expressions of emotion and empathy

Mirroring and counter-mirroring

Nearly 100 years ago Lipps (1907; in Hess and Blairy, 2001) suggested that when presented with a facial expression of emotion, *mirroring* (sometimes referred to in the literature as *imitation* or *mimicry*) leads – via a feedback process – to emotional contagion, which in turn facilitates emotional recognition. Rogers (1957) linked mirroring and empathy in the context of the actions of the counselor in response to a client, while in social psychology the nonverbal communication of emotion is generally assumed to occur primarily through the face (Lanzetta and Englis, 1989).

Mirroring is assumed to occur automatically as a response to observed facial expressions (Hoffman, 1984). This is consistent with the first two levels of Leventhal's model, with the innate central motor programs (first level) being triggered by facial stimuli, and experiences during development leading to a refinement of the responses, which remain automatic (second level). These two aspects of the perceptual-motor theory of emotion – that it is developmental, and that it proposes explicit mechanisms for the emergence of mirroring – set it apart from other prominent accounts, such as the Schematic, Propositional, Associative, and Analogical Representational Systems (SPAARS) model (Power and Dalgleish, 1997) or Teasdale's Interacting Cognitive Subsystems (Barnard and Teasdale, 1991).

Hess, Philippot, and Blairy (1999) reviewed the literature on mirroring and its role in the communication of emotion. They conclude that there is strong evidence (e.g., Dimberg, 1982; 1997; Lundqvist and Dimberg, 1995; Wallbott, 1991) that humans – both adults and infants – tend to mirror from a very early age. However, they also cite examples of counter-mirroring, i.e., expressing an alternative expression to that being presented. For example, Lanzetta and Englis (1989) found mirroring when participants believed they were collaborating but counter-mirroring when participants believed they were competing with one another. Hess and Blairy (2001) suggest that this implies mirroring may not be an automatic, reflex-like mechanism. However, it could be argued that the effect is simply

context dependent, remaining outside of conscious control nonetheless. A further alternative is that the initial, automatic response, when detected, can be suppressed at the conceptual level, at which point counter-mirroring is consciously adopted.

Evidence for a link between empathy and mirroring

In their review, Hess et al. (1999) found that the empirical support for the notion that mirroring facilitates emotion recognition is inconsistent. However, there remains the theoretical basis (from Lipps' model), linking mirroring to recognition through emotional contagion. In a study of their own (Hess and Blairy, 2001), they asked female participants to rate a short series of video clips of human faces expressing anger, sadness, disgust, and happiness. Participants were filmed during the task and a measure of emotional contagion was taken. Mirroring was measured using facial electromyography (EMG), with electrode placements chosen according to Fridlund and Cacioppo (1986). Mirroring was evident for all emotion classes, while limited support was found for emotional contagion of happiness and sadness. They found no relations between mirroring and contagion, nor between mirroring and recognition. However, Wixon and Laird (1981) did find evidence that recognition varies with mirroring. In their study, participants were asked to pose facial pain, inhibit, or act naturally (i.e. without instruction), in response to seeing another person receiving a shock. It should be noted that the emotional expression being looked at in the Wixon and Laird paper – that of pain – was not investigated in the Hess and Blairy study, which may account for the apparent inconsistency.

More recently a series of studies carried out by Marianne Sonnby-Borgström and her colleagues have demonstrated a link between empathy and mirroring (Sonnby-Borgström, 2002; Sonnby-Borgström, Jönsson and Svensson, 2003). In these studies participants were exposed to pictures of angry or happy faces at different exposure times. A significant difference in mirroring between high- and low-empathy participants was found at all exposure times. High-empathy participants showed significant mirroring, while low-empathy participants did not and in fact tended to smile in response to angry faces –

another example of counter-mirroring effects. In terms of Leventhal's model, it may be hypothesised that the central motor programmes (the first level, an automatic, reflex-like process) in less empathic individuals are absent or impaired in some way; alternatively, the developmental experiences of these individuals may be such that their conditioned responses differ from those of more empathic individuals (the second level, an automatic memory system, of the perceptual-motor theory of emotion). While Leventhal hypothesised that the central motor programs are unlikely to change much across the lifespan, this second level develops according to conditioning (Leventhal, 1984).

Measuring mirroring

Facial EMG is generally used to measure facial responsiveness in mirroring studies (e.g., Hess and Blairy, 2001; Sonnby-Borgström, 2002; Sonnby-Borgström et al., 2003). However, EMG training can be prohibitively expensive and time-consuming. An alternative is the Facial Action Coding System (FACS), a formal method of rating faces through observation, developed by Ekman and colleagues (Ekman and Friesen, 1978). FACS classifies facial movements as a series of "Action Units" (AUs), of which 46 account for changes in facial expression and 12 describe changes in gaze direction and head orientation. However, FACS can only be used effectively following extensive training, rendering it impractical for use by most clinicians. Moreover, Sonnby-Borgström (2002) cites a paper by Tassinary and Cacioppo (1992) as a justification for using EMG rather than FACS, stating that the mirroring reactions were expected to be weak and too indiscernible to be classified by observation. However, the Tassinary and Caccioppo paper does not specifically make such a claim about mirroring experiments. Rather, it states that FACS was shown to be insufficiently sensitive to detect viewers' emotional responses to television adverts (Graham, 1980) and also to emotionally charged slides of natural scenes (Cacioppo, Petty, Losch, and Kim, 1986). There is no reference to mirroring studies, nor to the possibility that emotional expressions could be identified without having to resort to a system such as FACS, which has clearly defined lower thresholds. Such a method for measuring mirroring can be found in a task by Wallbott (1991). During the first part of the

task, a group of twenty students were asked to judge the emotions expressed in a series of pictures of faces, while being secretly filmed. They were invited back two weeks later and asked to view their own footage and guess which emotions they were decoding. Wallbott found a correspondence between the two sets of judgement data that was above chance, supporting the notion that mirroring had occurred. The obvious argument against the validity of these findings (pointed out by Wallbott himself) is that during the second part of the task participants could remember the order in which they made their original responses. Against this, Wallbott points out two factors that should have diverted attention and minimised recollection: participants did a number of studies during the two-week interval, and a large number of slides (65) were used. Nonetheless, a more convincing way to conduct the analysis would have been to randomise the presentation of response footage for the second part of the task.

At present no study has looked at the facial responses of people who score highly on the PCL-R, or who have a personality disorder, to facial expressions of emotion. However, the work of Sonnby-Borgström and colleagues gives grounds to hypothesis that mirroring may occur to a lesser extent in this population, given both that they are known to score low on measures of empathy and that empathy has been found to correlate positively with mirroring. There is also evidence from one study that personality disorder is associated with reduced facial expressiveness in response to stimuli designed to induce either positive or negative emotional states (Renneberg, Heyn, Gebhard, and Bachmann, 2005). In the study, the stimuli were three-minute video clips taken from two films: a violent scene from "Cry Freedom" for the negative stimulus, and a slapstick sequence from "French Kiss" for the positive stimulus.

In keeping with the Government's current priorities for research into DSPD, it would be beneficial to investigate facial responsiveness to facial expressions of emotion for the following two reasons:

1. It may allow inferences to be made about mechanisms that contribute to the development of DSPD (e.g., impaired central motor programs or abnormal conditioned responses, or a combination of the two, in accordance with Leventhal's model).
2. Given that the treatment model currently being piloted at HMP Whitemoor includes emotional engagement as a core aspect, being able to incorporate a measure of mirroring into a test-retest battery may be useful as part of the treatment evaluation. Specifically, increases in facial mirroring in the DSPD offenders upon re-testing, relative to the control group, when taken in the context of a variety of other measures, could be interpreted as an indicator of treatment efficacy. In terms of Leventhal's model, this would constitute the development of new conditioned responses to the facial expressions of others.

This study investigates the facial responses to facial expressions of emotion of prisoners currently involved with the DSPD Programme. An adaptation of the facial affect recognition paradigm, pioneered by Ekman and his colleagues, is employed for the task, as this method has been used effectively in a number of mirroring studies, (e.g., Wallbott, 1991; Hess and Blairy, 2001). A brief history and critique of the paradigm follows, concluding this literature review.

Facial affect recognition tasks

There has been a debate for over 100 years in social psychology about whether facial expressions of emotion are universal or culture-specific, beginning with Charles Darwin's (1872) The Expression of the Emotions in Man and Animals (Ekman, 1999). Darwin believed that expressions of emotion may have originally served the function of enabling an organism to *cope* with the emotion and elements in the environment producing or maintaining the state. However, once these expressions were acquired they presented an additional, communicative function. Over the last 35 years the most frequently cited evidence in favour of the universality of facial expressions of emotion has come from the work of Paul Ekman and his colleagues (e.g., Ekman, Sorenson, and Friesen, 1969; Ekman

et al., 1987). The basic paradigm employed in much of this research involved presenting participants with a selection of photographs depicting prototypical facial expressions of emotion, and asking them to identify the particular emotion that was being displayed. In the earlier studies the expressions were posed by actors, but in a bid to provide greater ecological validity there have now been a number of studies using unposed expressions and even short video clips (Hess and Blairy, 2001), created by exposing participants to emotionally evocative stimuli and photographing or filming their responses.

In a summary of the evidence from facial affect recognition tasks, Ekman (1999) cites 31 groups from 21 literate countries who were shown photographs depicting happiness, anger, fear, sadness, disgust, and surprise. In every case, the majority in each group agreed about the photographs that showed happiness, sadness, and disgust. In 20 out of the 21 countries the majority agreed about the surprise faces, 19 out of 21 agreed about the fear faces, and 18 out of 21 about the anger faces. Where there was disagreement, the most frequent response given by participants was the same across all the atypical countries.

In recent years a number of studies have been conducted using the facial affect recognition paradigm with populations displaying psychopathic tendencies, or having particular personality disorders (e.g., Blair, Jones, Clark, and Smith, 1997; Blair, Colledge, Murray, and Mitchell, 2001; Blair, Mitchell, Peschardt, Colledge, Leonard, Shine, Murray, and Perret, 2004; and Kosson, Suchy, Mayer, and Libby, 2002). In the Blair et al. (1997) study psychopaths were found to display less autonomic activity than non-psychopaths while viewing sad or fearful faces but not while viewing more general (non-facial) threatening stimuli. In a study exposing participants to standardized emotional stimuli, again that were not facial expressions (Herpertz, Schwenger, Kunert, Lukas, Gretzer, Nutzmann, Schuerkens, and Sass, 2000) there were no differences in the psychophysiological reactions of patients with borderline personality disorder compared with patients with avoidant personality disorder and a non-clinical control group.

Blair et al. (2001) investigated the sensitivity of children with psychopathic tendencies to facial expressions. They found selective impairments in the recognition of sad and fearful

expressions, a finding later repeated in adults (Blair et al., 2004). Blair (2001) has argued that sad expressions constitute a human submission response, and that correct perception of them is necessary to inhibit aggression through what he has termed a Violence Inhibition Mechanism. A contrasting finding was reported by Kosson et al. (2002), who compared the responses of 34 incarcerated psychopaths to those of a group of 33 non-psychopaths. They found that psychopaths were less accurate than non-psychopaths at classifying facial affect, with a specific deficit in classifying disgust. And finally, Mikhailova, Vladimirova, Iznak, Tsusulkovskaya, and Sushko (1996) found that individuals with schizotypal personality disorder showed reduced recognition of *sad* and *happy* faces relative to a non-clinical control group. These studies demonstrate both that the facial affect recognition paradigm can be used with populations displaying psychopathic tendencies or having personality disorder diagnoses, and that significant differences related to affect have been found between these populations and non-clinical groups.

The classic facial affect recognition task has not gone without criticism, most notably from Russell (1994, 1995), who has argued against the claim for universality of expressions. Russell points out that in groups of participants who are isolated from so-called Western culture, agreement that smiles indicate something positive is high, but agreement about other expressions is low and can even reduce to chance when methodological artefacts are removed. These artefacts include a forced-choice response format, within-subject design, and pre-selected photographs of posed facial expressions. When these are altered, Russell demonstrates that less supportive or non-supportive results occur, while combining them may help to shape the results. Russell does concede, however, that there is likely to be an association between facial expressions and emotions, albeit one loose enough to permit alternative accounts to universality of the magnitude endorsed by Ekman. The present study takes some of Russell's criticisms into account and attempts to address them, opting to use un-posed expressions and a combination of free- and forced- selection response format.

Posed and un-posed facial expressions of emotion are similar, but there are detectable differences. For example, Smith, McHugo, and Lanzetta (1986) compared posed and imagery-induced expressions of happiness, sadness, and anger, and found that brow activity

present in the imagery-induced sad expressions was weak or absent in the posed ones. Forced *selection*, as opposed to forced *choice*, addresses Russell's criticism that the latter format portrays choices as mutually exclusive categories, by allowing participants to select more than one emotion for a single face (Russell, 1994). A completely free-selection response format addresses many of the limitations found when providing labels, including those outlined in the previous paragraph. However, these formats can be tedious for participants and analyzing open-ended participant responses can be time-consuming and ambiguous for experimenters.

Russell and others have also raised the objection that cross-cultural differences in emotional labelling mean that emotion words validated in North America, as is the case with much emotional recognition research, may not correspond to the emotional constructs of other cultures (Haidt and Keltner, 1999; Russell and Yik, 1996). In the present study this concern will be disregarded as the vast majority, if not all of the participants are expected to have a Western cultural background.

AIMS AND HYPOTHESES

The evidence of reduced mirroring in individuals who scored low on a measure of empathy generates the hypothesis that prisoners on the DSPD Programme (with elevated PCL-R scores, associated with low empathy) mirror less than the normal population also. The further hypothesis was made that prisoners from the general prison population would show impaired mirroring to a lesser extent, reflecting PCL-R scores falling somewhere between the normal population and DSPD. This research was intended to pilot a procedure to measure mirroring in a prison setting, and to investigate the hypothesis that psychopathic individuals show impaired mirroring. The procedure could also be developed to evaluate the outcomes of treatments that target such deficits.

METHOD

Participants

The study was conducted at Her Majesty's Prison (HMP) Whitemoor and at the University of Leeds. At HMP Whitemoor, all 49 male prisoners on the two treatment spurs of D Wing (the DSPD Unit) were invited by letter to participate as the experimental group. Prisoners on the Programme had chosen to be involved with it, having agreed to an assessment that includes a PCL-R. They were accepted onto the Programme subject to receiving a PCL-R score greater than 30, *or* a PCL-R score greater than 25 and at least one personality disorder diagnosis other than antisocial personality disorder, *or* a PCL-R score greater than 20 and at least two personality disorder diagnoses other than antisocial personality disorder.

Eleven prisoners from the DSPD Unit agreed to participate, with a mean age of 40.1 years (*SD* = 8.4, *Range* = 28.4 – 52.6). The mean number of personality disorder diagnoses was 2.9 (*SD* 1.5). Diagnoses included examples of all but one (*dependent* personality disorder) of the ten personality disorders featured in DSM-IV-TR. Fig. 2 shows the frequencies of the different diagnoses, indicating that the *antisocial* and *borderline* subtypes were the most common.

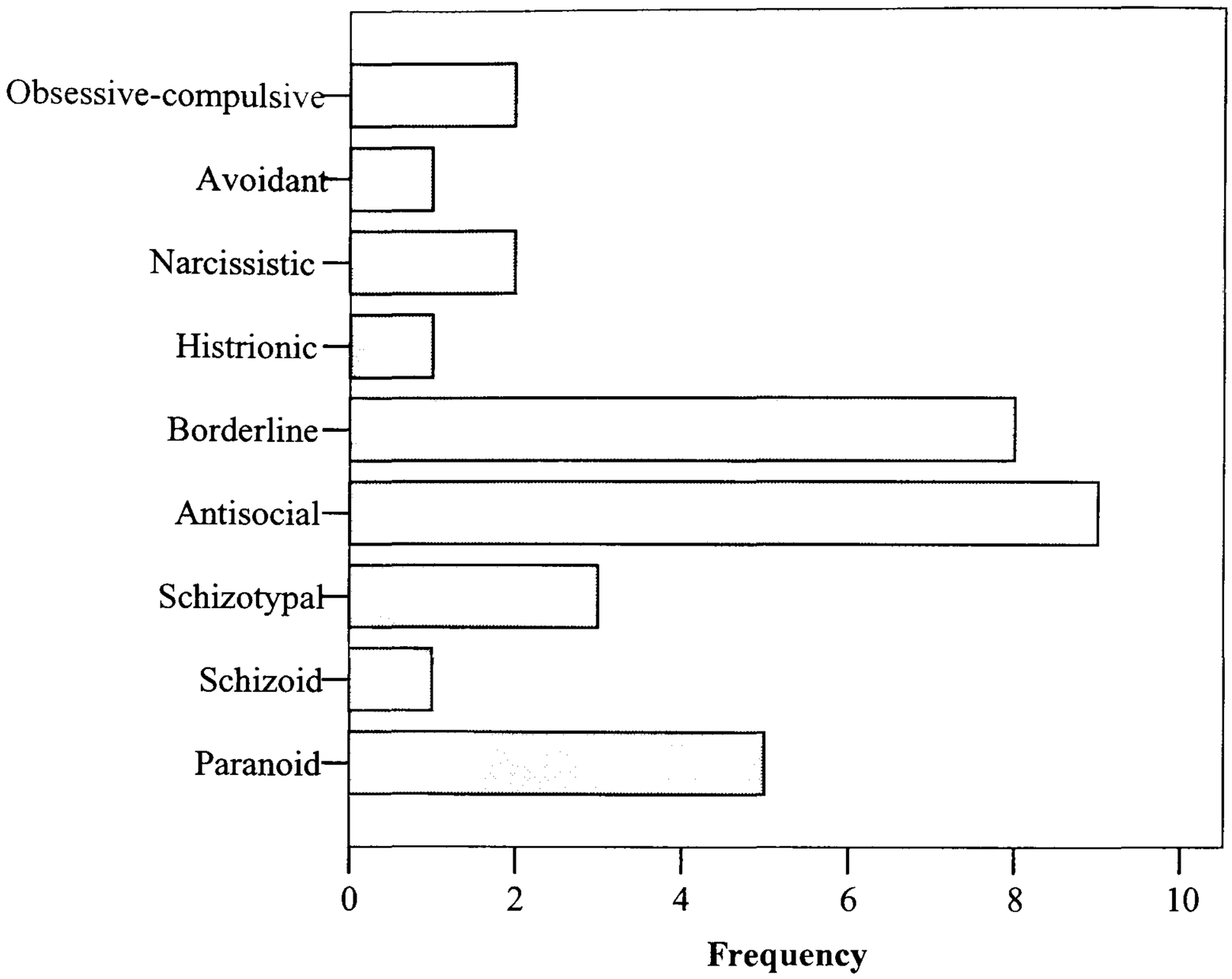

Fig. 2: Frequency of different personality disorder diagnoses in the DSPD sample.

The mean PCL-R score for ten of the eleven participants was 28.4 (*SD* = 3.6, *Range* = 24-33). One participant had a PCL-R score of 24, i.e., below the frequently used cutoff of 25. Because of the small group size, and taking into consideration the Harris et al. (1994) finding that a score as low as 19-20 can actually optimize the hit-rate for psychopathy, the data from this participant was not excluded from the analysis. The eleventh participant did not have a PCL-R score, but had received a Psychopathy Checklist - Screening Version (PCL-SV) score of 18. A PCL-SV score does not convert directly to a PCL-R score – unlike the PCL-R, it contains no criminal items – but scores correlate highly ($r = 0.8$) with the PCL-R; the authors suggest that a score of 18 or over is strongly indicative of psychopathy (Hart, Cox, and Hare, 1995).

Two control groups were recruited for the study. The first was a group of nine predominantly Category A and B prisoners – none of whom were involved with the DSPD Programme – housed on A Wing at HMP Whitemoor. Letters of invitation were sent to all 111 male prisoners on the wing, 41 of whom were accommodated on a voluntary drug

testing spur, while the remaining 71 prisoners were accommodated in regular cells. The initial response was low: only two prisoners agreed to participate. The author therefore visited the wing to provide prisoners with the opportunity to have any questions about the study answered. A major concern for some individuals centred on the need to be filmed during the task. Reassurances were given about the confidentiality of the footage, and that it would be destroyed at the end of the study. Seven additional prisoners agreed to participate, making a group total of nine participants. The mean age of this group was 38.9 years (*SD* = 11.2, *Range* = 25.4 – 61.2). Personality disorder diagnoses and PCL-R scores were not available; however, participants were asked whether they were aware of having received any such diagnoses or a PCL-R score, with it being made clear that they were under no obligation to share this information. Only one participant stated that he had received a personality disorder diagnosis (antisocial).

The third control group was recruited through advertisements on the University of Leeds Reporter website and posters put up around the University, and comprised a combination of male students and staff working at the University. In total ten people agreed to participate (mean age 31.1 years, SD 13.5). Informed consent was gained from all participants, and the study was approved by Northern and Yorkshire Multi-centre Research Ethics Committee (see Appendices A-H for ethical approval, letters of invitation, participant information sheets and informed consent forms).

Apparatus and materials

A Toshiba Satellite 4070CDT laptop, with 1024x768 pixel screen resolution was used to display the video clips using Windows Media Player. Clips were edited using Adobe Premiere. A video camera was needed in order to record participants' faces during the facial affect recognition task. It was not possible to use the same camera for the prison groups as for the University group: security clearance could not be obtained to take a camera into the prison, and it would have been similarly prohibited to return the prison's own camera, should it have been borrowed for use with the University group. A Panasonic M-10 VHS camcorder was therefore used with the prison groups, and a Sony GR-DVL 150

digital video camera was used with the University group. Both were mounted on adjustable tripods.

Stimuli

Stimuli were taken from the Video Database of Moving Faces and People, created by Alice O'Toole and her colleagues at the University of Texas, Dallas (O'Toole, Harms, Snow, Hurst, Pappas, Ayyad, and Abdi, 2005). Participants in the database project were filmed as they viewed a ten-minute video, which contained scenes from various movies and television programs intended to elicit different emotions. Camera angle and lighting conditions were controlled throughout, allowing for consistency of presentation. The database includes five-second video clips of dynamic, unposed emotional expressions for 284 participants that have not been formally normed by the authors. In total there are 2,150 clips in the database.

A subset of clips was sought from the database for this research, comprising five clips – at least two from each gender – and depicting each of the six basic emotions (Ekman, Friesen, and Ellsworth, 1972): *anger, disgust, fear, happiness, sadness,* and *surprise*. An additional ten clips - five from each gender – were also sought to be used as *neutral* stimuli. The *neutral* clips were intended to serve as a counterbalance to the preponderance of negative emotions depicted by the other clips. Four of the clips for each emotion, plus eight of the *neutral* clips, were to be used in the main task, while the remaining ten clips would be used in a practice task to familiarise participants with the procedure.

Within the constraints of this research it was deemed impractical to have the database rated in its entirety by multiple raters. The author therefore viewed all 2,150 clips in order to filter and discard those that were unlikely to be of use, either because they clearly showed blends of emotions or because they were particularly weak examples of specific emotions. Following this initial exercise, a number of useable clips were found for each of the emotions: *anger* (38 clips), *disgust* (43 clips), *fear* (44 clips), *sadness* (30 clips), and

surprise (43 clips). There were many more examples of *happiness* (95 clips) and *neutral* (406 clips). The *mean* number of clips for the other five emotions (*Mean* ~ 40) was therefore used to determine the number of *happiness* and *neutral* clips selected for rating, i.e., 40 *happiness* clips were randomly selected and, in order to reflect the balance of clips that were sought for the main task, 80 (twice as many) *neutral* clips were selected. This resulted in 322 clips to be used in the rating task.

During the initial filtering of the clips it was noted that there were large variations in the duration of expressions. Some expressions lasted for most of the full 5 seconds, while others were much shorter, sometimes lasting only a fraction of a second. In order to balance the duration of expression across clips – thereby ensuring that this variable was held constant for each emotional class – it was decided at this stage to edit the clips so that they would freeze and hold at full expression. This was a difficult decision, as editing the clips in this way was clearly at the expense of a degree of ecological validity; however, it was reasoned that the resulting set of clips would nonetheless constitute spontaneous expressions of emotion. Moreover, pausing clips at full expression plausibly creates stimuli that are no more artificial than the still images used in the vast majority of emotion recognition studies.

Adobe Premiere was used to edit the clips. As far as was possible, clips were edited so that the onset of expression commenced 12 image frames, or approximately 0.5 seconds, after the start of the clip (in a small minority of clips the onset of expression commenced slightly earlier than this). The clips were then set to pause at the peak of the expression and hold as a still image until the end of the 5-second duration. These start and end points were ascertained by careful frame-by-frame viewing of the clips, through which it was possible to locate the points at which groups of facial muscles would first begin to move from one arrangement and second settle into another. The mean number of frames taken for a face to change from starting expression to final expression was 18.38 ($SD = 9.62$). This figure was used to determine the number of frames into each clip that the *neutral* clips were then paused in order to balance them with the rest of the set, i.e. the *neutral* clips were edited to pause 30 (12 plus 18) frames, or 1.2 seconds, after the start of the clip.

The full set of 322 edited clips was viewed and rated independently by seven people, recruited from among the author's peers and colleagues (four males and three females; mean age 27.1 years, *SD* = 5.9). Raters were asked to label the emotion or emotions that they perceived in each clip, from a selection of *anger, disgust, fear, happiness, sadness, surprise, neutral,* and *other* (to be specified). Each response was rated on a seven-point scale from one (*weak*) to seven (*strong*) (see Appendix I for instructions). Clips were presented on a laptop in a random order using Windows Media Player. Raters were able to pause the presentation after each clip was presented while they considered their answer, before starting it again. The instructions were placed alongside the computer to act as a reminder of the labelling options, and answers were given verbally and recorded by the author. This response method was employed in order to enable the volunteer raters to progress more quickly through the clips – which nonetheless took up to 90 minutes to rate – as they were free to view the next clip while the author recorded their previous response.

The initial criterion applied when selecting clips for the main task was agreement between at least five of the seven raters (71 per cent) on the presence of a particular emotion. This procedure provided a sufficient number of clips depicting *anger, disgust, happiness, surprised*, *neutral* and *sadness* (female only). There were no male *sadness* clips, nor *fear* clips of either gender that met this criterion. It was necessary therefore to use clips with lower rates of agreement for these expressions (4 for the female *fear* and male *sadness* expressions and 3 for the male *fear* expressions; see Table 1). A further complication was presented with *fear*, as all of the clips meeting this criterion were actually more likely to be rated as *surprise* clips, either in combination with *fear* or another emotion, or as purely *surprise*. The clips selected for this emotion were therefore labelled as blends of *fear* and *surprise*.

	Minimum rater agreement criterion					
	3		4		5	
Emotion	Male	Female	Male	Female	Male	Female
Anger	15	25	7	10	5*	5*
Disgust	19	46	14	39	12*	32*
Fear/surprise	5*	17	0	6*	0	0
Happiness	24	24	20	21	20*	20*
Sadness	5	38	2*	27	0	12*
Surprise	21	68	20	59	18*	55*
Neutral	35	39	32	32	24*	15*

Table 1: Number of clips in each emotion class depicting each gender, with different minimum rater agreement criteria applied. *Criterion.

From the available clips (indicated with an asterisk in Table 1), the two male and two female clips with the highest mean strength rating within each emotion class were selected for use in the main task, while the third highest strength rating clips were selected for use in the practice clips (three male and three female clips). A random selection of five male and five female clips labelled *neutral* were also selected. The resulting stimulus set is summarized in Table 2, indicating mean strength ratings and standard deviations as calculated using only the scores of raters who actually labelled the emotion. As is clear from the table, there is variability across emotions. The male *sadness* clips received particularly weak ratings, while the female *disgust* clips were rated much more strongly. It would have been possible to balance this out somewhat by selecting weaker clips from those emotions for which more clips were available, thereby weakening the stimulus set overall; however, it was reasoned that it would be better to use the strongest clips available for each emotion, with the hypothesis that these clips were more likely to lead to mirroring by participants during the main task. Consequently, in the analysis for the main task, each emotional class was analysed separately; for example, mirroring scores are calculated for *anger*, *disgust*, *fear/surprise*, etc., but no overall mirroring score is given.

Emotion	Number of clips	Mean strength rating (*SD*) Male	Female
Anger	5	3.4 (0.4)	3.6 (0.28)
Disgust	5	3.85 (0.02)	6.33 (0.22)
Fear/ Surprise	5	4.54 (0.66)/ 4.08 (0.12)	3.77 (0.41)/ 3.25 (0.57)
Happiness	5	4.67 (0.3)	5.22 (0.3)
Sadness	5	2.38 (0.53)	4.89 (0.6)
Surprise	5	4.44 (1.29)	4.97 (0.66)
Neutral	10	N/A	N/A

Table 2: Number of clips selected for main task, including practice clips.

Design

The main task involved a 3 (Group: DSPD, prison control, University control) x 7 (Emotional expression: *anger, disgust, fear/surprise, happiness, sadness, surprise, neutral*) mixed design. Group was a between-participant variable and emotional class was a within-participant variable.

Procedure

A facial affect recognition task was carried out in rooms on the two wings at HMP Whitemoor and in a room at the University of Leeds. The laptop was set up on a table to run the task, and Participant Information Sheets, Task Instructions and response booklets were placed next to the laptop. A video camera was mounted on a tripod in the corner of the room and left on standby. This was used to film participants' faces during the task, to be rated by independent raters at a later date. Participants were informed that they were to be filmed, and that the recordings would be viewed by independent raters.

Efforts were made to ensure the different environments were as similar as possible, particularly with respect to lighting, by always placing the table before a window and directly beneath a strip light on the ceiling. Fig. 1 indicates the features of the rooms that were held constant, with the centimetre measurements referring to the distance from one corner of the table to the lens of the camera – which was also positioned at a height of 86cm. There was some small variation between participants on this measure, as the camera had to be slightly repositioned to accommodate the different heights and seating postures of participants.

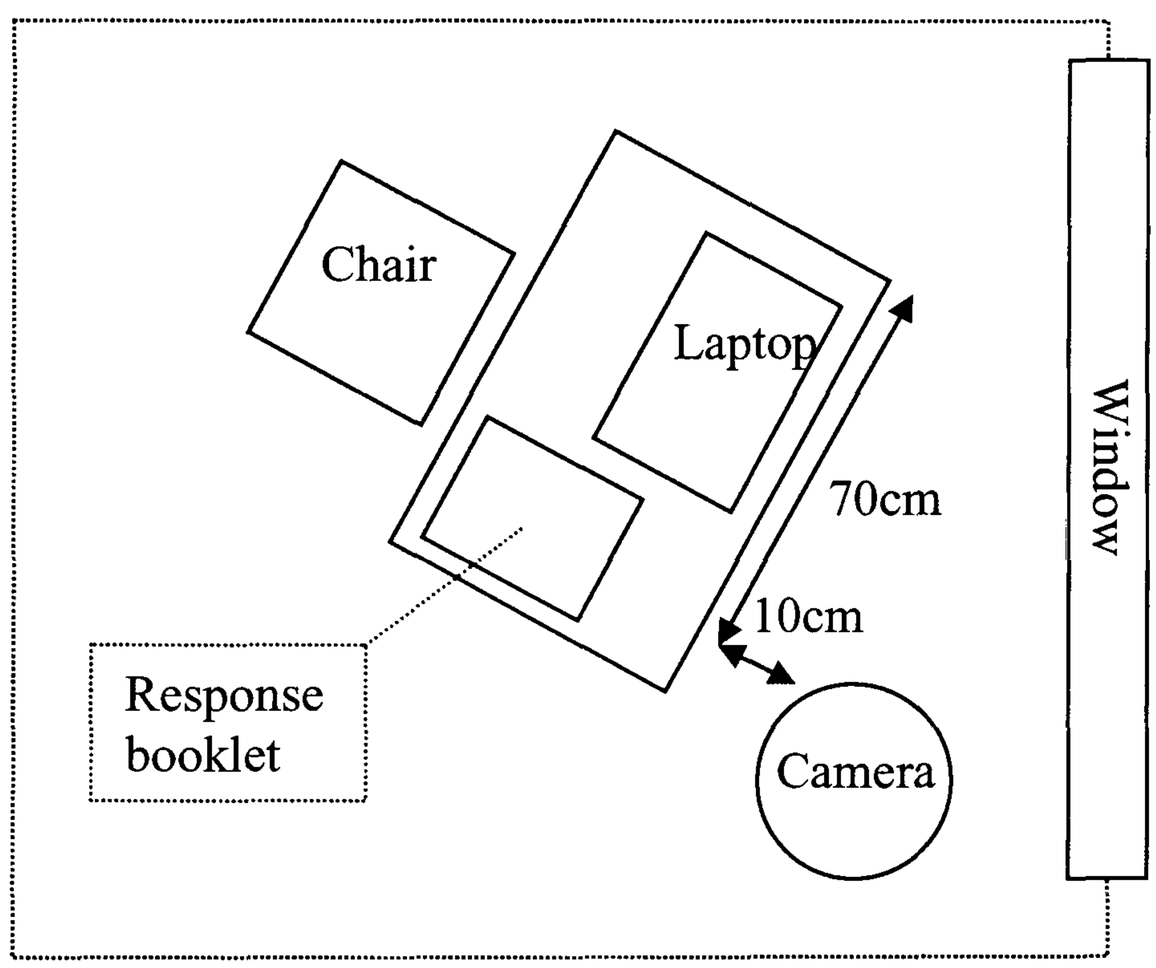

Fig. 3: Illustration of the features of the rooms that were held constant across settings.

Participants were asked to watch each clip all the way through, and to wait for it to disappear before indicating their response from a selection of *anger, disgust, fear, happiness, sadness, surprise, neutral* and *other*. If *other* was selected, participants were also asked to write the word they felt best described the emotion in the clip. Each response was rated on a seven-point scale from one (*weak*) to seven (*strong*), and participants were able to assign more than one emotion to each clip (see Appendix J for instructions). Prior to commencing the practice trials they were shown an example of the countdown screen that would appear between clips, and that also displayed the number of the clip that had just

been shown (in order to ensure that the correct answer was filled in on the response booklet).

For the practice task the eight selected clips were presented in a random order. The task served both to familiarize participants with the procedure and to enable the author to start and position the video camera correctly. The participant was informed of the moment the camera was switched on, which was done immediately before presentation of the practice clips. It also provided a period of time for participants to become accustomed to the presence of the camera. Immediately following this participants were given the opportunity to have any remaining questions answered before completing the main task, which consisted of 64 clips (each of the 32 main task clips presented twice), shown in a random order, with an opportunity for a short break halfway through.

During the practice task the time between clips gradually reduced, from an initial 24 seconds for the first two clips, to 19 seconds for the third and fourth clips and finally to 14 seconds for the remaining four clips. Each clip was preceded by two audio beeps, providing a two-second count-in to the clip. The audio tones were to be used later for editing the footage of participants' facial responses during the task, but served also as a prompt to participants that a clip was about to appear on-screen. The practice task took approximately four minutes to complete and the main task lasted approximately 20 minutes.

Rating

The final task prior to analysis was to take a measure of participants' facial responses during the main task. This was done by showing the video footage – comprising a portrait view of each participant as they completed the task – to two independent raters, who were asked to label the emotions being expressed by participants. One rater was a trainee forensic psychologist and the other was a psychologist in clinical training. Neither of the two raters worked directly with either of the prisoner groups. However, one of the raters – who held an honorary contract with the prison – had conducted some research on D Wing

that included brief meetings with some of the prisoners. Neither rater had any special training in the decoding of facial expressions, as was the case in the successful Wallbott (1991) study.

Before the rating was conducted, the footage was segmented and randomly ordered into five sets of clips. Each set was intended to be used in a task that could be completed by a rater in a single sitting. In order to achieve this, the footage was first converted to *digital video* (DV) files. Subsequently, it was segmented using Adobe Premiere to produce 64, five-second clips for each participant (or 1 920 clips in total). These clips corresponded to the periods during which stimuli were being viewed during the main task (the audio beeps were used to locate the periods when participants were viewing stimuli).

It was impractical to have all 1 920 clips rated in a single sitting, as this would take around eight hours. It was also assumed that raters would learn to read the faces of individual participants over a number of trials. It was therefore preferable for all of the clips from any one participant to be encountered within a single rating session, and for blocks of clips to have a duration that would not induce fatigue in the raters. In order to achieve this, clips from subgroups of six participants (two selected at random from each group) were compiled to form five blocks, each one taking approximately 90 minutes to rate. Because the original group numbers were not equal, for the fifth and final set it was necessary to include three participants from the DSPD group, one from the prison control group, and two from the University group. In summary, five blocks were created, with each one comprising the 64 main task clips for six participants, totalling 384 clips per block. The clips within each block were presented in a (different) random order to each rater.

The rating task was similar in form to the familiar facial affect recognition paradigm. It differed primarily in the information that was given about the nature of the clips: raters were informed that the people in the clips had themselves been viewing facial expressions of emotion that included *anger*, *disgust*, *fear*, *happiness*, *sadness*, *surprise*, and *neutral*. They were also advised that sometimes people mirror the facial expressions that they see. They were then asked to guess the emotion that was being viewed for each clip. Raters

selected only one emotion for each clip, from a selection of *anger, disgust, fear, happiness, sadness, surprise,* and *neutral* (see Appendix K for instructions). It had been noted by the author at the editing stage that the vast majority of the facial responses of participants during the task were of very low intensity. Raters were therefore not asked to provide a strength score. Clips were presented to the raters one set at a time, using Windows Media Player. A screen was displayed after each clip for 20 seconds, displaying the number of the clip that had just been shown. Raters were able to skip forward to the next clip once they had marked their answer in the response booklet, and short breaks could be taken after every 64 clips. Each of the five sets took approximately 90 minutes to complete.

RESULTS

Main analysis

Interrater reliability

Unweighted Cohen's kappa scores for interrater reliability were low (mean $\kappa = 0.28$, $SD = 0.15$, *Range* = 0.12 – 0.6). Kappas for each set of clips corresponding to a single emotional class in the stimulus set are summarised in Table 3, which indicates low mean values for all classes (*Range* = 0.19 – 0.38). The data from each rater were therefore analysed separately, rather than only trials on which both raters agreed being considered.

	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral
Mean κ	0.27	0.36	0.38	0.29	0.27	0.31	0.19
SD	0.1	0.11	0.15	0.14	0.1	0.14	0.18
Range	0.12 – 0.42	0.23 – 0.56	0.20 – 0.6	0.05 – 0.46	0.11 – 0.43	0.16 – 0.6	-0.12 – 0.51

Table 3: Mean Cohen's kappa scores for each emotional class (based on the stimulus set classifications).

The low kappas may in part be explained by a low prevalence of observable mirroring: this suggestion is supported by the fact that the majority of clips were labelled neutral in the

rating task (63.4% of clips by Rater 1 and 78.1% of clips by Rater 2; see Table 4). Such a result produces an asymmetrical imbalance in the marginal totals of the kappa table – in this case, a 7 x 7 table – which can generate low values for κ despite a higher level of agreement (Kraemer, 1979). Simple agreement was 67%, much higher than the kappas but low nonetheless.

Rater	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral
1	6.4	2.6	2.6	9.9	7.6	7.6	63.4
2	2.9	4.0	1.7	6.7	2.8	3.9	78.1

Table 4: Percentage of clips mapped onto each emotional class.

Mirroring

Raw mirroring scores

In order to measure the extent to which mirroring occurred, rater responses were compared with the "true" identity of the clip that was being viewed (as determined by the original rating of the stimulus set). The percentage of trials (for each emotional class of stimuli, i.e. *anger, disgust, fear/surprise, happiness, sadness, and surprise*) on which there was a match between rater response and clip identity provided raw mirroring scores for each participant for each emotional class. A response to clips corresponding to *fear/surprise* clips was classed as mirroring if the rater entered either *fear* or *surprise*.

Kolmogorov-Smirnov D (normality) tests and Levene's homogeneity of variance tests were carried out on the raw mirroring scores with *group* as a between-subjects factor. These are summarised in Tables 5 and 6. All mirroring variables were non-normal, a fact that remained unchanged when the data were transformed using either logarithmic or square root transformations. Furthermore, mirroring of *happiness* as scored by Rater 1, plus mirroring of both *anger* and *surprise* as scored by Rater 2, lacked homogeneity of variance.

Mirroring variable	Group	Kolmogorov-Smirnov *D*	N	Sig.	Levene *F*	Sig.
Anger	DSPD	0.27	11	0.03	2.89	0.07
	Prison control	0.41	9	0.00		
	University control	0.41	10	0.00		
Disgust	DSPD	0.41	11	0.00	1.18	0.32
	Prison control	0.52	9	0.00		
	University control	0.48	10	0.00		
Fear/ surprise	DSPD	0.23	11	0.10	0.20	0.82
	Prison control	0.23	9	0.20		
	University control	0.36	10	0.00		
Happiness	DSPD	0.34	11	0.00	7.71	0.00
	Prison control	0.32	9	0.01		
	University control	0.23	10	0.15		
Sadness	DSPD	0.31	11	0.01	0.43	0.66
	Prison control	0.28	9	0.04		
	University control	0.31	10	0.01		
Surprise	DSPD	0.26	11	0.04	2.46	0.10
	Prison control	0.26	9	0.09		
	University control	0.33	10	0.00		

Table 5: K-S *D*-scores and Levene's test statistics for raw mirroring as scored by Rater 1. For the Levene's tests, $k - 1 = 2$ and $N - k = 27$ degrees of freedom.

Mirroring variable	Group	Kolmogorov-Smirnov			Levene	
		D	*N*	*Sig.*	*F*	*Sig.*
Anger	DSPD	0.39	11	0.00	6.04	0.01
	Prison control	0.40	9	0.00		
	University control	0.48	10	0.00		
Disgust	DSPD	0.34	11	0.00	0.81	0.46
	Prison control	0.52	9	0.00		
	University control	0.32	10	0.00		
Fear/ surprise	DSPD	0.28	11	0.01	1.56	0.23
	Prison control	0.31	9	0.02		
	University control	0.36	10	0.00		
Happiness	DSPD	0.43	11	0.00	0.21	0.81
	Prison control	0.43	9	0.00		
	University control	0.37	10	0.00		
Sadness	DSPD	0.48	11	0.00	1.15	0.30
	Prison control*	-	-	-	-	-
	University control	0.40	10	0.00		
Surprise	DSPD	0.35	11	0.00	4.24	0.02
	Prison control	0.47	9	0.00		
	University control	0.48	10	0.00		

Table 6: K-S *D*-scores and Levene's test statistics for raw mirroring as scored by Rater 2. For the Levene's tests, $k - 1 = 2$ and $N - k = 27$ degrees of freedom. *There was no mirroring of sadness in the prison control group; this variable was therefore constant and was excluded from these tests.

Descriptive statistics for raw mirroring data are summarised in Tables 7 (Rater 1) and 8 (Rater 2). These show that the majority of variables were positively skewed and leptokurtic. As the data were not normally distributed, median scores are shown instead of means as a

more meaningful measure of central tendency. These medians were generally low, with a range of 0 – 25% (Rater 1), and 0 – 12.5% (Rater 2).

Mirroring variable	Group	*Skewness*	*Kurtosis*	*Median*	*Min.*	*Max.*
Anger	DSPD	1.24	0.87	12.50	0	63
	Prison control	0.86	-1.71	0.00	0	13
	University control	1.64	1.06	0.00	0	50
Disgust	DSPD	2.41	6.06	0.00	0	38
	Prison control	3.00	9.00	0.00	0	50
	University control	1.78	1.41	0.00	0	13
Fear/ surprise	DSPD	1.01	1.01	12.50	0	63
	Prison control	0.56	0.19	25.00	0	50
	University control	1.05	-0.39	0.00	0	38
Happiness	DSPD	0.65	-1.55	0.00	0	25
	Prison control	2.27	5.66	0.00	0	50
	University control	0.59	-1.20	25.00	0	88
Sadness	DSPD	1.30	0.60	0.00	0	50
	Prison control	2.11	4.71	0.00	0	63
	University control	0.47	-1.81	6.25	0	25
Surprise	DSPD	1.81	3.91	12.50	0	75
	Prison control	0.50	-0.01	12.50	0	38
	University control	0.00	-2.57	6.25	0	13

Table 7: Descriptive statistics for raw mirroring variables as scored by Rater 1.

Mirroring variable	Group	*Skewness*	*Kurtosis*	*Median*	*Min.*	*Max.*
Anger	DSPD	0.91	-1.27	0.00	0	25
	Prison control	1.19	-0.45	0.00	0	25
	University control	1.78	1.41	0.00	0	13
Disgust	DSPD	2.79	8.33	0.00	0	63
	Prison control	3.00	9.00	0.00	0	38
	University control	0.13	0.18	12.50	0	25
Fear/ surprise	DSPD	0.34	-1.70	12.50	0	38
	Prison control	2.06	4.68	12.50	0	63
	University control	1.18	0.57	0.00	0	25
Happiness	DSPD	1.83	2.45	0.00	0	38
	Prison control	2.80	7.97	0.00	0	63
	University control	2.45	6.78	12.50	0	75
Sadness	DSPD	2.42	5.51	0.00	0	25
	Prison control*	-	-	0.00	0	0
	University control	2.27	5.36	0.00	0	38
Surprise	DSPD	2.05	4.19	0.00	0	50
	Prison control	1.62	0.73	0.00	0	13
	University control	1.78	1.41	0.00	0	13

Table 8: Descriptive statistics for raw mirroring variables as scored by Rater 2. *There was no mirroring of sadness in the prison control group; values for skewness and kurtosis are therefore not given.

Parametric tests are robust to mild violations of core assumptions (e.g., Howell, 1999); however, because more than one assumption was violated on a number of the mirroring variables, statistical tests were applied conservatively throughout, i.e. non-parametric tests were used. Kruskal-Wallis *H* tests were carried out, testing for differences on mirroring

scores between groups. None of the group differences reached statistical significance at the 5% level. However, in order to investigate the possibility of significant pairwise differences, a more liberal level of 15% was applied. For five variables, $p < 0.15$ (see Table 9). Pairwise analyses were carried out on these (Mann-Whitney *U* tests), in order to determine whether pairs of groups differed significantly in mirroring. For these tests, the more robust 5% level was used to determine significant group differences (see Table 10).

		Rater 1		Rater 2	
Mirroring variable	Group	*H*	*Sig.*	*H*	*Sig.*
Anger	DSPD Prison control University control	4.82	0.09*	1.23	0.54
Disgust	DSPD Prison control University control	0.60	0.83	4.95	0.08*
Fear/ surprise	DSPD Prison control University control	2.65	0.27	1.55	0.47
Happiness	DSPD Prison control University control	3.78	0.15*	4.07	0.13*
Sadness	DSPD Prison control University control	0.02	0.99	2.97	0.23
Surprise	DSPD Prison control University control	3.97	0.15*	1.25	0.57

Table 9: Results of Kruskal-Wallis *H* tests (*d. f.* = 2) carried out on raw mirroring, shown separately for Rater 1 and Rater 2. *$p < 0.15$.

Rater	Mirroring variable	Groups	Mann-Whitney *U*	*Sig.*	*r*
1	Anger	DSPD – Prison controls	24.00	0.03**	-0.38
		DSPD – University controls	34.50	0.07	-0.28
		Prison controls – University controls	43.50	0.44	-0.03
1	Happiness	DSPD – Prison controls	47.00	0.43	-0.04
		DSPD – University controls	32.00	0.05*	-0.31
		Prison controls – University controls	26.50	0.06	-0.29
1	Surprise	DSPD – Prison controls	48.50	0.49	-0.01
		DSPD – University controls	32.50	0.06	-0.31
		Prison controls – University controls	25.00	0.06	-0.32
2	Disgust	DSPD – Prison controls	38.00	0.18	-0.21
		DSPD – University controls	38.50	0.12	-0.23
		Prison controls – University controls	22.00	0.01*	-0.39
2	Happiness	DSPD – Prison controls	47.50	0.49	-0.04
		DSPD – University controls	35.00	0.07	-0.28
		Prison controls – University controls	25.00	0.04*	-0.33

Table 10: Results of Mann-Whitney *U* tests carried out on differences in raw mirroring scores between pairs of groups. 1-tailed significance levels are reported. *$p < 0.05$. **$p < 0.05$; however, this difference is in the opposite direction to that predicted and is therefore non-significant.

Taking into account the status of this research as a pilot study, and in the interests of identifying potentially important group differences that could be explored in future research, in addition to the liberal *p*-values, familywise corrections (e.g., the Bonferroni procedure) were not applied throughout the analysis. Those differences that approached significance at the initial level of analysis were therefore investigated further with pairwise tests.

One-tailed significance levels are reported, in accordance with the hypothesis that mirroring will be greatest in the university control group and lowest in the DSPD group. Effect sizes were calculated for all variables, using Pearson's correlation coefficient *r*. For mirroring of happiness as scored by Rater 1, university controls (*Mdn* = 25) were found to mirror more than participants in the DSPD group (*Mdn* = 0), with a medium effect size ($U = 32$, $p = 0.048$, $r = -0.31$). For mirroring of both *disgust* and *happiness* as scored by Rater 2, university controls (*Mdn* = 12.5 for each variable) were found to mirror significantly more than prison controls (*Mdn* = 0 for each variable), with medium effect sizes ($U = 22$, $p = 0.01$, $r = -0.39$; $U = 25$, $p = 0.04$, $r = -0.33$, respectively). Mirroring of *anger* as scored by Rater 1 was higher in the DSPD group (*Mdn* = 12.5) than in the prison controls (*Mdn* = 0), with a medium effect size ($U = 24$, $p = 0.03$, $r = -0.38$); however, this difference was not in the predicted direction and therefore precludes rejection of the null hypothesis for mirroring of *anger*.

Controlling for baseline facial expressions

The above method generated raw scores for mirroring that did not take into account potential individual or group differences in participants' baseline facial expressions. It was important to control for this potential source of variance because, for example, it could be argued that any group differences in the mirroring of *anger* expressions are a reflection of the natural tendency of members of the DSPD group to appear more angry than the other groups (irrespective of the particular stimuli they are viewing). There is indirect empirical evidence for this suggestion in reports of elevated trait anger levels in offender populations and forensic patients (Chemtob, Novaco, Hamada, and Gross, 1997; Watt and Howells, 1999).

Rater responses for clips when participants were viewing *neutral* stimuli were used to establish base rates of expressions. It was assumed that footage of participants viewing *neutral* clips would give an indication of their neutral expression. Taking *anger* as an

example, for each participant the percentage of clips corresponding to *neutral* stimuli that a rater labelled *anger* was subtracted from that participant's raw mirroring score for *anger*. The same procedure was applied to the remaining five emotional classes.

Kolmogorov-Smirnov D tests and Levene's tests were again carried out to measure normality and homogeneity of variance respectively, with composite mirroring scores as dependent variables and *group* as a between-participants factor. Tables 11 and 12 show K-S and Levene statistics for Rater 1 and Rater 2 respectively. Descriptive data are shown in Tables 13 and 14. As was the case with raw mirroring scores, no variables met both normality and homogeneity of variance assumptions for parametric testing, with a tendency to be positively skewed and leptokurtic. Group medians were lower than for raw mirroring, as would be expected following a subtraction from participants' scores, falling within the range 0 – 15.63% (Rater 1) and 0 – 9.38% (Rater 2).

Composite mirroring variable	Group	Kolmogorov-Smirnov *D*	*N*	*Sig.*	Levene *F*	*Sig.*
Anger	DSPD	0.29	11	0.01	1.11	0.35
	Prison control	0.34	9	0.00		
	University control	0.46	10	0.00		
Disgust	DSPD	0.37	11	0.00	1.89	0.17
	Prison control	0.48	9	0.00		
	University control	0.52	10	0.00		
Fear/ surprise	DSPD	0.21	11	0.17	0.34	0.72
	Prison control	0.19	9	0.20		
	University control	0.31	10	0.01		
Happiness	DSPD	0.23	11	0.11	6.11	0.01
	Prison control	0.21	9	0.20		
	University control	0.23	10	0.15		
Sadness	DSPD	0.31	11	0.01	0.80	0.46
	Prison control	0.24	9	0.16		
	University control	0.27	10	0.03		
Surprise	DSPD	0.31	11	0.00	2.97	0.07
	Prison control	0.23	9	0.20		
	University control	0.34	10	0.00		

Table 11: K-S *D*-scores and Levene's homogeneity test statistics for composite mirroring as scored by Rater 1. For the Levene's tests, $k - 1 = 2$ and $N - k = 27$ degrees of freedom.

Composite mirroring variable	Group	Kolmogorov-Smirnov *D*	*N*	*Sig.*	Levene *F*	*Sig.*
Anger	DSPD	0.36	11	0.00	5.08	0.01
	Prison control	0.37	9	0.00		
	University control	0.35	10	0.00		
Disgust	DSPD	0.24	11	0.08	2.26	0.12
	Prison control	0.52	9	0.00		
	University control	0.27	10	0.05		
Fear/ Surprise	DSPD	0.17	11	0.20	2.88	0.07
	Prison control	0.26	9	0.07		
	University control	0.43	10	0.00		
Happiness	DSPD	0.41	11	0.00	0.15	0.86
	Prison control	0.40	9	0.00		
	University control	0.33	10	0.00		
Sadness	DSPD	0.48	11	0.00	8.29	0.00
	Prison control	0.52	9	0.00		
	University control	0.43	10	0.00		
Surprise	DSPD	0.26	11	0.04	4.67	0.02
	Prison control	0.33	9	0.01		
	University control	0.40	10	0.00		

Table 12: K-S *D*-scores and Levene's homogeneity test statistics for composite mirroring as scored by Rater 2. For the Levene's tests, $k - 1 = 2$ and $N - k = 27$ degrees of freedom.

Composite mirroring variable	Group	*Skewness*	*Kurtosis*	*Median*	*Min.*	*Max.*
Anger	DSPD	1.74	4.43	12.50	-6.25	50.00
	Prison control	0.66	-0.42	0.00	-6.25	12.50
	University control	2.04	3.55	0.00	-6.25	43.75
Disgust	DSPD	1.80	4.44	0.00	-12.50	37.50
	Prison control	2.55	7.45	0.00	-12.50	43.75
	University control	3.16	10.00	0.00	0.00	12.50
Fear/ surprise	DSPD	1.39	2.09	6.25	-6.25	50.00
	Prison control	0.38	-0.56	12.50	-12.50	43.75
	University control	1.45	1.57	0.00	-12.50	37.50
Happiness	DSPD	-0.08	0.00	0.00	-18.75	25.00
	Prison control	1.03	1.84	0.00	-18.75	31.25
	University control	0.88	-0.49	15.63	-6.25	87.50
Sadness	DSPD	1.07	0.27	0.00	-12.50	31.25
	Prison control	-0.35	0.93	0.00	-18.75	25.00
	University control	1.15	0.82	3.13	-6.25	25.00
Surprise	DSPD	1.98	3.88	6.25	-6.25	68.75
	Prison control	1.51	2.33	6.25	0.00	37.50
	University control	0.81	1.24	0.00	-6.25	12.50

Table 13: Descriptive statistics for composite mirroring variables as scored by Rater 1.

Composite mirroring variable	Group	*Skewness*	*Kurtosis*	*Median*	*Min.*	*Max.*
Anger	DSPD	0.78	-0.99	0.00	-6.25	25.00
	Prison control	1.02	-0.29	0.00	-6.25	25.00
	University control	1.02	1.83	0.00	-6.25	12.50
Disgust	DSPD	1.82	4.88	0.00	-18.75	50.00
	Prison control	3.00	9.00	0.00	0.00	18.75
	University control	1.33	2.21	6.25	0.00	25.00
Fear/ surprise	DSPD	0.59	-0.85	6.25	-12.50	37.50
	Prison control	1.82	3.64	6.25	-6.25	56.25
	University control	1.59	2.13	0.00	-6.25	18.75
Happiness	DSPD	1.71	2.16	0.00	-6.25	37.50
	Prison control	2.39	6.06	0.00	-6.25	43.75
	University control	2.68	7.80	9.38	0.00	75.00
Sadness	DSPD	2.42	5.51	0.00	0.00	12.50
	Prison control	-3.00	9.00	0.00	-6.25	0.00
	University control	1.26	-0.07	0.00	0.00	18.75
Surprise	DSPD	1.48	2.83	0.00	-18.75	50.00
	Prison control	0.94	1.35	0.00	-6.25	12.50
	University control	0.00	4.50	0.00	-12.50	12.50

Table 14: Descriptive statistics for composite mirroring variables as scored by Rater 2.

Kruskal-Wallis *H* tests were carried out, comparing composite mirroring scores between groups. None of the group differences reached significance at the 5% level. The more liberal level of 15% was therefore applied again in order to investigate the possibility that

there were significant pairwise differences. For four emotions, differences were significant at the 15% level (see Table 15).

Composite mirroring variable	Group	Rater 1 *H*	Rater 1 *Sig.*	Rater 2 *H*	Rater 2 *Sig.*
Anger	DSPD Prison control University control	2.86	0.25	1.64	0.45
Disgust	DSPD Prison control University control	0.58	0.85	4.00	0.14*
Fear/ surprise	DSPD Prison control University control	1.92	0.39	1.13	0.58
Happiness	DSPD Prison control University control	4.10	0.13*	4.20	0.12*
Sadness	DSPD Prison control University control	1.15	0.57	4.52	0.12*
Surprise	DSPD Prison control University control	2.28	0.33	0.23	0.89

Table 15: Results of Kruskal-Wallis *H* tests (*d. f.* = 2) carried out on composite mirroring, shown separately for Rater 1 and Rater 2. *$p < 0.15$.

Pairwise analyses were carried out on those variables for which $p < 0.15$ (Mann-Whitney U tests), in order to determine whether pairs of groups differed significantly in composite mirroring. For these tests, the more robust 5% level was used to determine significant group effects (see Table 16). For composite mirroring of *happiness* as scored by Rater 1, university controls (*Mdn* = 15.63) were found to mirror more than prison controls (*Mdn* = 0), with a medium effect size ($U = 22.5$, $p = 0.03$, $r = -0.34$). For Rater 2, university controls mirrored significantly more *happiness* (*Mdn* = 9.38) and *disgust* (*Mdn* = 6.25) clips than prison controls (*Mdn* = 0 for each variable), with medium effect sizes ($U = 23$, $p = 0.03$, $r = -0.34$; $U = 21$, $p = 0.01$, $r = -0.4$).

Rater	Mirroring variable	Groups	Mann-Whitney *U*	Sig.	*r*
1	Happiness	DSPD – Prison controls	35.50	0.14	-0.20
		DSPD – University controls	38.00	0.11	-0.23
		Prison controls – University controls	22.50	0.03*	-0.34
2	Disgust	DSPD – Prison controls	49.00	0.50	-0.01
		DSPD – University controls	37.00	0.10	-0.24
		Prison controls – University controls	21.00	0.01*	-0.40
2	Happiness	DSPD – Prison controls	43.00	0.33	-0.10
		DSPD – University controls	35.00	0.07	-0.27
		Prison controls – University controls	23.00	0.03*	-0.34
2	Sadness	DSPD – Prison controls	36.00	0.14	-0.30
		DSPD – University controls	46.50	0.22	-0.15
		Prison controls – University controls	28.00	0.07	-0.36

Table 16: Results of Mann-Whitney U tests carried out on differences in composite mirroring scores between pairs of groups. One-tailed significance levels are reported. *$p < 0.05$.

Supplementary analyses

Recognition

Recoding responses

Before assessing rates of recognition of the stimuli among participants, the response data were recoded in order to overcome the difficulties inherent in analysing (partially) open-ended responses. There were two stages to this procedure: first, responses that participants entered in the *other* category were reclassified either as one of the six basic emotions (used to define the stimulus set), or as *neutral*. In the second stage, multiple responses (i.e., emotional classes) entered on a single trial were collapsed into one, dominant response.

The affective lexicon developed by Andrew Ortony and colleagues (Ortony, Clore, and Foss, 1987; Clore, Foss, and Ortony, 1987) was used to facilitate the first stage. This lexicon comprises 585 words taken from the literature on emotion. In the first of two papers, the researchers report a taxonomy of these words (Ortony, Clore, and Foss, 1987), derived using componential analysis (Goodenough, 1956). The rationally (as opposed to empirically) derived taxonomy is shown in Fig. 10. In the diagram, psychological conditions are shown in ellipses, while rectangles enclose the features that differentiate them. The following eight categories are contained in the taxonomy, collapsed into four broad classes (illustrated in the diagram by the four rectangles with thick borders):

- Subjective Evaluations and Objective Descriptions (External Conditions class)
- Physical and Bodily States (Physical and Bodily States class)
- Affective States, Affective-Behavioural Conditions, Affective-Cognitive Conditions (Affective Conditions class)
- Behavioural-Cognitive Conditions and Cognitive Conditions (Cognitive Conditions class)

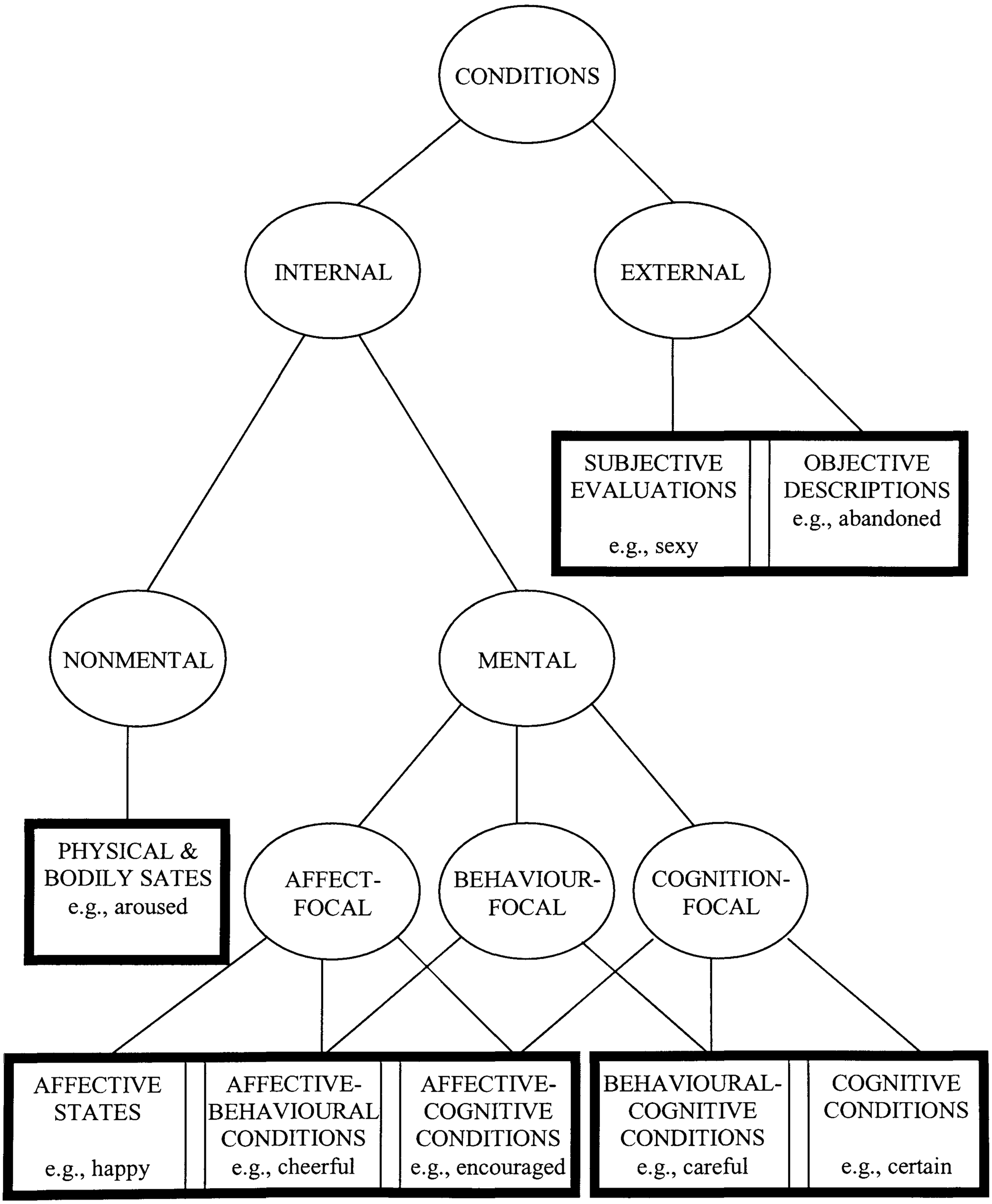

Fig. 4: Taxonomy of psychological conditions (based on Ortony, Clore, and Foss, 1987). Ellipses depict psychological conditions and rectangles show the features that differentiate them.

Ortony and colleagues proposed that the best examples of emotions are those words that fall into the Affective Conditions class, i.e., psychological conditions that are internal, mental, and affect-focal. The second paper provided empirical support for the theoretical taxonomy (Clore, Ortony, and Foss, 1987). In the study, participants rated their confidence that words from the lexicon referred to an emotion in the two contexts of *feeling something* and *being something*. The researchers predicted that the words they had pre-experimentally (or rationally) classified in the Affective Conditions class, i.e., emotions, would score highly in both contexts, that Cognitive Conditions would score moderately in both, that scores in both contexts would be low for Physical and Bodily States, and that for External Conditions scores would be high in the *feeling* context and low in the *being* context. Using discriminant analysis, the predictions were correct for the majority of words in each class. However, a minority of words that were rationally classified as emotions are empirically classified as one of the other three classes, and vice versa.

Returning to this study, 24 participants (80%) made use of the *other* category, contributing a total of 116 different words or phrases. These responses were classed as an emotion if they appeared in the Affective Conditions class of the affective lexicon for either the rational classification, or the empirical classification, or both. Responses that did not appear in the Affective Conditions class in either context (rational or empirical) were reclassified as *neutral*. A thesaurus was used to identify synonyms for the resulting 30 responses that were classed as emotions (Compact Oxford Dictionary Thesaurus; OUP, 2002). For 28 of these, synonyms were found either directly (e.g., *annoyed = angry*), or through two steps (e.g., *amused = pleased = happy*), making it possible to reclassify each response as one of the six emotional classes (*anger, disgust, fear, happiness, sadness,* or *surprise*) or, in the single case of *apathy*, as *neutral*. One response, *gobsmacked*, did not appear in the (American) lexicon; the dictionary definition of this word confirmed it is an informal British term for "utterly astonished" (OUP, 2002). *Astonished*, which did appear in the lexicon as an empirically classified Affective Condition, was then transformed to *surprise* as a direct synonym. Finally, *violent* was empirically classified as an Affective Condition; however, there were no synonyms in the thesaurus used that made it possible to map this word to one of the seven basic emotion classifications (including *neutral*). *Violent* occurred

on only one occasion, and was accompanied by a response of *anger* on the same trial. On this basis, *violent* was assumed to constitute an elaboration of *anger*, rather than an additional emotion. The mapping for all words in the *other* category can be found in Appendix L.

The second stage of recoding the recognition data involved collapsing the sometimes multiple responses made by participants on a single trial, into one, dominant emotional class. In the majority of cases, this was done by simply selecting the response with the highest strength rating. The small minority of trials on which two or more emotional classes were rated with equal strength (<2% of trials) were treated as missing data, rather than arbitrarily selecting a single emotion as predominant.

Rates of recognition

Using the recoded data, recognition rates were calculated for each participant for each emotional class (as percentage correct). Responses of either *fear* or *surprise* to the *fear/surprise* clips were designated correct. Recognition rates varied substantially between emotional classes: *Mean* = 62.1%, *SD* = 18.2, *Range* = 36.4 – 93.2%, in the DSPD group; *Mean* = 73.4%, *SD* = 10.5, *Range* = 19.4 – 95.8%, in the prison control group; and *Mean* = 72.3%, *SD* = 12.5, *Range* = 50 - 95%, in the university control group.

Tests of normality and homogeneity of variance were carried out on the recognition scores (see Table 17). While not all variables were normally distributed, variance was similar between groups for all recognition variables. ANOVAs tend to be robust to violations of a single assumption (Lindman, 1974); one-way between-participant ANOVAs were therefore carried out on the data to compare group mean recognition scores, with *group* as the between-participants factor and recognition scores as the dependent variables. The results of these tests are summarised in Table 18, which shows that there is a significant effect of *group* on recognition of *anger*, $F(2, 27) = 4.7$, $p = 0.02$, $\omega = 0.45$.

Recognition variable	Group	Kolmogorov-Smirnov			Levene	
		D	d.f.	Sig.	*F*	Sig.
Anger	DSPD	0.31	11	0.00*	1.13	0.34
	Prison control	0.20	9	0.20		
	University control	0.20	10	0.20		
Disgust	DSPD	0.19	11	0.20	0.05	0.95
	Prison control	0.24	9	0.15		
	University control	0.17	10	0.20		
Fear/ surprise	DSPD	0.25	11	0.05*	0.84	0.44
	Prison control	0.26	9	0.08		
	University control	0.19	10	0.20		
Happiness	DSPD	0.40	11	0.00*	0.66	0.52
	Prison control	0.41	9	0.00*		
	University control	0.46	10	0.00*		
Sadness	DSPD	0.21	11	0.20	0.50	0.61
	Prison control	0.29	9	0.03*		
	University control	0.21	10	0.20		
Surprise	DSPD	0.31	11	0.00*	1.29	0.29
	Prison control	0.37	9	0.00*		
	University control	0.21	10	0.20		
Neutral	DSPD	0.18	11	0.20	0.48	0.62
	Prison control	0.29	9	0.03*		
	University control	0.22	10	0.19		

Table 17: K-S *D*-scores and Levene's homogeneity test statistics for recognition scores. For the Levene's tests, $k - 1 = 2$ and $N - k = 27$ degrees of freedom. $*p < 0.05$, i.e., a non-normal variable.

Recognition variable	Group	*Mean*	*SD*	ANOVA *F*	*Sig.*
Anger	DSPD	36.36	25.28	4.71	0.02*
	Prison control	19.44	17.80		
	University control	50.00	20.41		
Disgust	DSPD	73.86	21.25	0.07	0.93
	Prison control	77.78	27.80		
	University control	75.00	22.05		
Fear/ surprise	DSPD	67.05	26.97	2.09	0.14
	Prison control	87.50	13.98		
	University control	76.25	22.40		
Happiness	DSPD	93.18	15.17	0.13	0.88
	Prison control	95.83	6.25		
	University control	95.00	12.08		
Sadness	DSPD	47.73	21.52	1.86	0.18
	Prison control	63.89	18.16		
	University control	53.75	15.65		
Surprise	DSPD	73.86	30.85	0.93	0.41
	Prison control	88.89	21.14		
	University control	82.50	19.72		
Neutral	DSPD	52.27	33.10	2.12	0.14
	Prison control	77.08	27.06		
	University control	73.13	27.01		

Table 18: ANOVAs ($d.f. = 27, 2$) carried out on recognition scores. $k - 1 = 2$ and $N - k = 27$ degrees of freedom. *$p < 0.05$.

Tukey's HSD was applied to anger scores in order to identify pairwise differences; this test has good statistical power and tightly controls for Type I errors when sample sizes are

approximately equal and population variances are likely to be similar (Field, 2005). Normal controls were found to have significantly higher scores for recognition of *anger* than prison controls ($p = 0.01$); there were no significant differences between the DSPD group and either of the two control groups. Fig. 11 shows an error bar chart, which clearly illustrates the significant group differences on recognition of anger, plus the overlapping confidence intervals of the DSPD group with the two control groups on this variable.

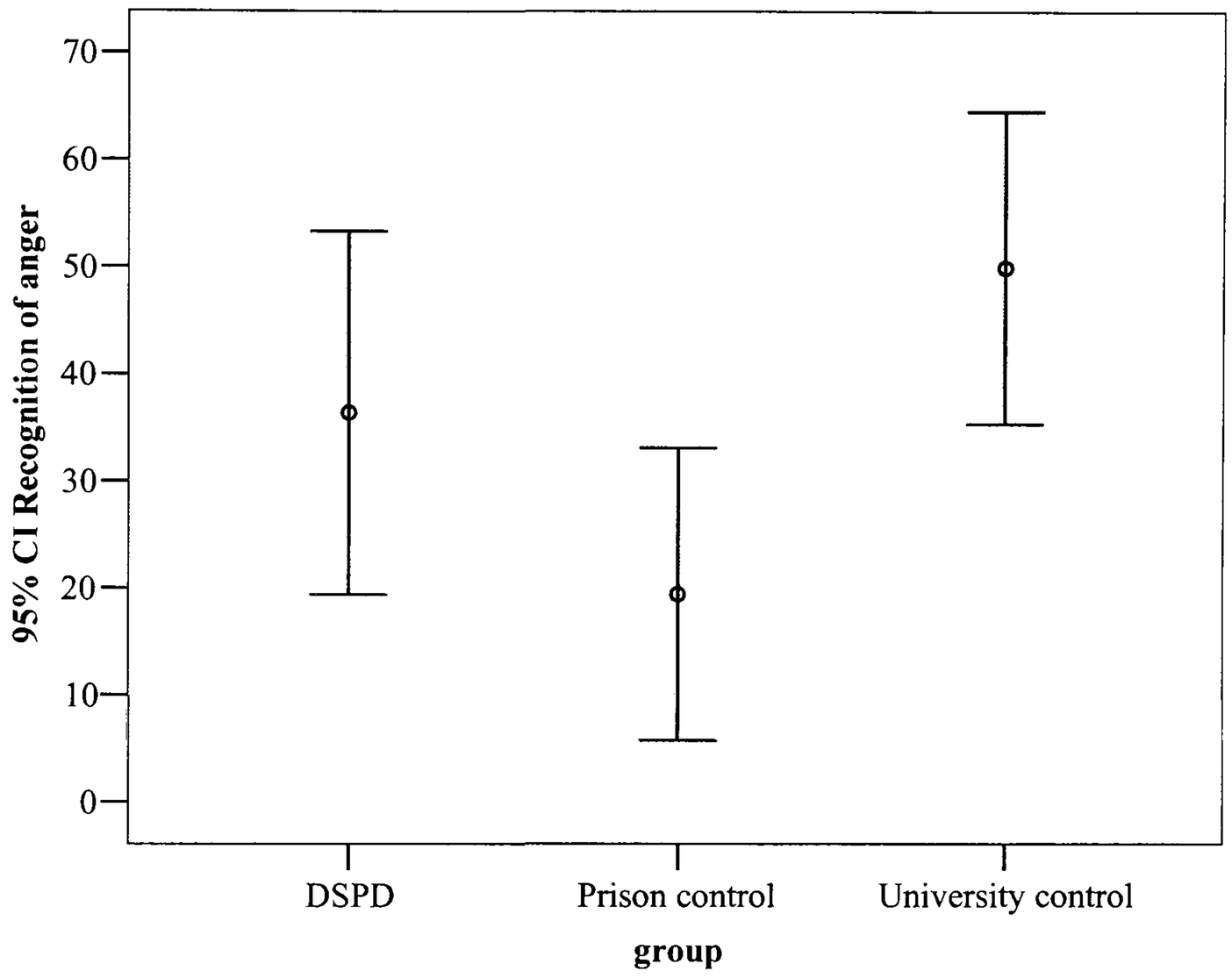

Fig. 5: Error bar chart showing the group differences on recognition of anger with 95% Confidence Interval bars.

Recognition versus mirroring

Spearman's rank order correlation coefficients were computed to explore the associations between composite mirroring and recognition scores. Data for each group were analysed separately, with coefficients calculated for each emotional class. For example, *composite mirroring of anger* was correlated with *recognition of anger*, and so on for the other emotional classes. The results are summarised in Table 19, which shows that the

coefficients were non-significant (for Rater 1, *Mean r* = -0.1 (*SD* = 0.29), *Mean p* = 0.49 (*SD* = 0.25); for Rater 2, *Mean r* = -0.11 (*SD* = 0.24), *Mean p* = 0.61 (*SD* = 0.29)). The only finding that came close to significance was the association between mirroring and recognition of sadness (Rater 2). However, contrary to hypothesis, this relationship was negative, i.e., lower mirroring was associated with increased recognition (*r* = -0.55, *p* = 0.08).

		Rater 1		Rater 2	
Group	Emotion	*r*	*Sig.*	*r*	*Sig.*
DSPD	Anger	0.33	0.32	0.14	0.68
	Disgust	-0.29	0.39	0.01	0.98
	Fear/surprise	0.03	0.92	0.04	0.91
	Happiness	-0.10	0.78	-0.05	0.89
	Sadness	-0.01	0.97	-0.55	0.08
	Surprise	-0.25	0.46	-0.36	0.28
Prison control	Anger	0.40	0.29	-0.02	0.96
	Disgust	-0.19	0.63	-0.21	0.59
	Fear/surprise	0.08	0.83	-0.20	0.61
	Happiness	0.14	0.72	0.35	0.35
	Sadness	-0.37	0.33	-0.22	0.58
	Surprise	-0.44	0.23	0.04	0.92
University control	Anger	-0.30	0.39	-0.50	0.14
	Disgust	-0.24	0.51	-0.28	0.44
	Fear/surprise	-0.32	0.37	0.03	0.93
	Happiness	-0.38	0.28	-0.26	0.46
	Sadness	-0.37	0.29	-0.20	0.58
	Surprise	0.51	0.13	0.24	0.50

Table 19: Spearman's rank order correlation coefficients for composite mirroring and recognition.

Emotional sensitivity

A measure of emotional sensitivity was computed by calculating the mean strength ratings given by participants for each emotional class and overall. Responses of *neutral*, for which participants were not asked to give a strength rating, were given a score of one. On trials with two or more strength ratings, a number of possible procedures were considered to convert the trial to an overall strength rating. Taking the example of a face rated as seven on the *anger* scale and one on the *disgust* scale, the ratings could simply be summed; however, this would result in the clip having a strength higher than seven – which is conceptualised as *the most expressive a face can be*. A second alternative would be to take an average; however, this would diminish the perceived strength of the example given by nearly half. It was therefore assumed that the highest (or joint highest) rating on a trial gives the most meaningful description of its overall perceived strength.

Tests of normality and homogeneity of variance on emotional sensitivity scores are summarised in Table 17. The picture was mixed, with three variables (sensitivity to *anger*, *disgust,* and *neutral* clips) showing significant deviations from the normal distribution for one or more groups ($p < 0.05$), and two variables (sensitivity to *sadness* and *neutral* clips) lacking homogeneity of variance ($F(2, 27) = 6.35$, $p < 0.01$; $F(2, 27) = 3.85$, $p < 0.05$). Median sensitivity scores fell within the ranges 1.50 – 5.13 (DSPD group), 1.06 – 5.13 (prison controls), and 1.19-5.32 (university controls). Statistical tests were applied conservatively, i.e., non-parametric tests were used, across all variables when testing for group differences. The results of Kruskal-Wallis *H* tests, with emotional sensitivity scores as dependent variables and *group* as the between-participants factor, are summarised in Table 18. None of the group differences reached significance at the 5% level; however, in order to identify possible pairwise differences, a more liberal criterion ($p < 0.2$) was applied. For *sensitivity to anger*, $p = 0.15$, and for *sensitivity to neutral clips*, $p = 0.18$.

Sensitivity variable	Group	Kolmogorov-Smirnov			Levene	
		D	*N.*	*Sig.*	*F*	*Sig.*
Anger	DSPD	0.26	11	0.03	0.74	0.48
	Prison control	0.12	9	0.20		
	University control	0.22	10	0.19		
Disgust	DSPD	0.15	11	0.20	0.03	0.97
	Prison control	0.18	9	0.20		
	University control	0.28	10	0.03		
Fear/ surprise	DSPD	0.11	11	0.20	3.19	0.06
	Prison control	0.24	9	0.13		
	University control	0.20	10	0.20		
Happiness	DSPD	0.21	11	0.18	1.18	0.32
	Prison control	0.14	9	0.20		
	University control	0.14	10	0.20		
Sadness	DSPD	0.20	11	0.20	6.35	0.01
	Prison control	0.17	9	0.20		
	University control	0.16	10	0.20		
Surprise	DSPD	0.19	11	0.20	1.03	0.37
	Prison control	0.12	9	0.20		
	University control	0.15	10	0.20		
Neutral	DSPD	0.25	11	0.05	3.85	0.03
	Prison control	0.33	9	0.01		
	University control	0.29	10	0.02		

Table 20: K-S *D*-scores and Levene test statistics for emotional sensitivity scores. For the Levene's tests, $k-1=2$ and $N-k=27$ degrees of freedom.

Sensitivity variable	Group	*Median*	*Min.*	*Max.*	Kruskal-Wallis *H*	Sig.
Anger	DSPD	2.88	2.13	6.13	3.74	0.15*
	Prison control	2.38	1.25	4.13		
	University control	3.87	1.38	5.25		
Disgust	DSPD	4.67	2.88	5.80	2.36	0.31
	Prison control	5.13	3.38	6.25		
	University control	5.32	3.29	6.00		
Fear/ surprise	DSPD	5.13	2.88	6.75	0.80	0.67
	Prison control	5.13	2.88	5.86		
	University control	4.58	4.17	6.00		
Happiness	DSPD	5.00	3.25	6.75	1.16	0.56
	Prison control	4.88	3.50	6.25		
	University control	5.00	3.17	6.13		
Sadness	DSPD	3.29	2.38	5.75	0.51	0.78
	Prison control	3.38	2.25	4.71		
	University control	3.46	2.43	4.13		
Surprise	DSPD	4.63	3.63	6.80	1.86	0.39
	Prison control	4.25	3.13	5.88		
	University control	4.75	3.75	6.25		
Neutral	DSPD	1.50	1.00	5.14	3.40	0.18*
	Prison control	1.06	1.00	1.93		
	University control	1.19	1.00	3.44		

Table 21: Results of Kruskal-Wallis *H* tests (*d.f.* = 2) carried out on emotional sensitivity scores.

*$p < 0.2$.

Two-tailed pairwise tests were carried out on those variables that came closest to significance ($p < 0.20$); however, none of these reached significance at the 5% level (see Table 22).

Sensitivity variable	Groups	Mann-Whitney *U*	Sig.	*r*
Anger	DSPD – Prison controls	28.50	0.11	-0.29
	DSPD – University controls	49.00	0.71	-0.08
	Prison controls – University controls	24.00	0.09	-0.31
Neutral	DSPD – Prison controls	29.00	0.13	-0.29
	DSPD – University controls	47.50	0.61	-0.10
	Prison controls – University controls	26.00	0.13	-0.29

Table 22: Results of Mann-Whitney tests carried out on differences in emotional sensitivity scores between pairs of groups. Two-tailed significance levels are reported.

Number of emotional classes selected

The number of emotional classes selected per clip was compared across groups, with the hypothesis that participants in the university control group would make more selections per clip than the prison controls, while the DSPD group would score lowest on this measure. This hypothesis was based on the evidence that psychopathy is associated with impairments in affect recognition (Blair et al., 2001; Kosson et al., 2002), and on the assumption of elevated PCL-R scores in the prison control group, and even greater scores in the DSPD group. It was reasoned that the mechanism underlying the impaired recognition shown in previous studies, may manifest in this research as a reduction in the tendency to identify emotional *blends*, i.e., make multiple responses on a single trial.

Fig. 5 shows a boxplot of the number of selections per clip for each group. While data fell between one and two selections per clip for all three groups, it is apparent that there was a greater range covered by the university controls.

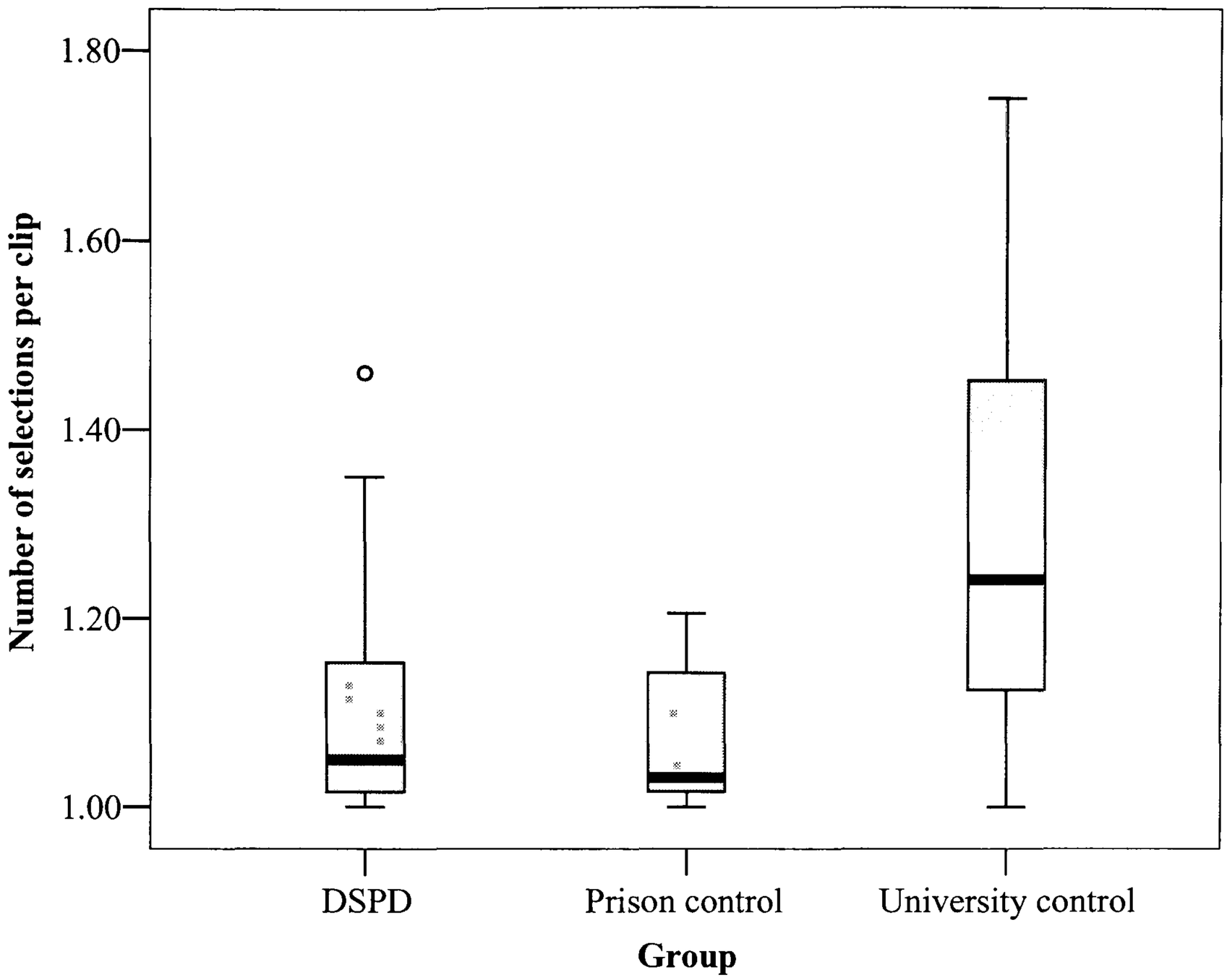

Fig. 6: Boxplot showing the number of selections per clip for each group. The circle above DSPD indicates a score with a *z*-value of 1.45, which is not classed as an outlier (Tabachnick and Fidell, 2001).

Kolmogorov Smirnov *D* tests indicated that the data for both DSPD and prison controls were non-normally distributed ($D(11) = 0.27$, $p = 0.03$; $D(9) = 0.29$, $p = 0.03$), and Levene's test statistic showed that variances were heterogeneous ($F(2, 27) = 5.17$, $p = 0.01$). A Kruskal Wallis *H* test comparing number of emotions selected per clip across groups did not reach significance at the 5% level ($H(2) = 5.4$, $p = 0.07$). However, as this result was close to significance, one-tailed pairwise tests were administered. As predicted, university controls ($Mdn = 1.24$) made more selections per clip than prison controls ($Mdn = 1.03$) and participants in the DSPD group ($Mdn = 1.05$), with medium effect sizes ($U = 19.5$, $p = 0.02$, $r = -0.38$; $U = 28$, $p = 0.03$, $r = -0.35$, respectively). However the difference between the two prison samples was non-significant.

DISCUSSION

Summary of findings

The results of this preliminary study do not show convincing evidence for reduced mirroring of facial expressions of emotion in prison populations. When more liberal *p*-values were applied, there were tentative suggestions that a larger study, or a more sensitive procedure, may reveal group differences. The clearest suggestions (significant only at the 15% level) were that mirroring of *disgust* and *happiness* occurred more frequently among university controls than prison controls. Differences between university controls and the DSPD group on these measures were also (nonsignificantly) in the predicted direction.

Applying Leventhal's perceptual-motor theory, it may be hypothesized that the inconsistent pairing of specific facial expressions and feeling states during development could disrupt the normal activation of the innate central motor programmes. This may occur with inconsistent parenting that is perceived by the developing child as frightened or frightening, such that both fear and security are evoked by the same source. Such parenting is thought to play a causal role in the development of a cluster of interpersonal behaviours labelled disorganized attachment style. There is a clear association between disorganized attachment and sexual and violent offending in a personality-disordered population (Van Ijzendoorn, Feldbrugge, Derks, De Ruiter, Verhagen, Philipse, Vand der Staak, and Riksen-Walraven, 1997). It is therefore probable that developmental primary caregiver relationships for both prison samples were less consistent than those for the university controls.

An alternative explanation for the possible reduced or absent mirroring in prison populations is that the necessary central motor programmes are not present, or else are not activated, in some individuals congenitally. According to this interpretation, for such individuals mirroring would not occur irrespective of the quality of interpersonal relationships during development. As with many other human attributes, teasing apart the relative contributions of genes and environment to the development of mirroring would be far from straightforward; however, it could be hypothesized that a larger environmental contribution would make it feasible that mirroring could be improved using a form of

interpersonal skills training, targeting the deficit directly. Conversely, if some individuals lack the necessary central motor programmes altogether, it should not be possible to normalize true mirroring.

Because of the uneven strengths among the stimulus clips, it is difficult to draw any firm conclusions about the two particular emotional classes (*disgust* and *happiness*) for which group differences approached significance. However, it is worth noting that during the initial ratings of stimuli for the task, the *disgust* and *happiness* clips received higher mean strength scores than any of the other emotional classes. This fact lends support to the hypothesis that stronger facial expressions are more likely to induce mirroring, and may indicate that the other clips used in this study were not sufficiently strong. Using a stimulus set comprising strong examples of all emotional classes may therefore have revealed more global mirroring deficits in the prison groups.

Contrary to previous studies that have investigated emotional recognition in samples with high PCL-R scores, deficits were not found in the recognition of *disgust* (Kosson et al., 2002), *fear*, or *sadness* (Blair et al., 1997; Blair et al., 2001). The only significant group difference was in the recognition of *anger*, which was found to be higher in the university controls than in the prison controls. However, for all three groups, recognition of *anger* was lower than for any other emotional class; moreover, in the initial rating of the stimuli the *anger* clips received the lowest mean strength ratings. An implication of this finding is that weaker stimuli may be required in order to reveal group differences in recognition of emotion. Taken together with the earlier suggestion that stronger clips are needed in order to observe group differences in mirroring, it may be difficult to measure both mirroring and recognition in a DSPD population within the same task. A possible solution would be to use a combination of extreme and moderate facial stimuli, balanced across emotional classes.

The lack of an association between mirroring and recognition, despite the theoretical rationale for there being a positive relationship, is in line with the findings of previous research (e.g., Blairy et al., 1999). The one relationship that approached statistical

significance in this part of the analysis (the negative association between mirroring and recognition of *sadness* in the DSPD group (Rater 2)) was in the opposite direction to that hypothesized. The simplest interpretation is that this association approached significance as a consequence of the number of statistical tests carried out without making familywise error corrections. Because of the exploratory nature of this research, such procedures were not applied, making it more likely that small group differences would be observed, while greatly increasing the probability of Type I errors occurring. Alternatively, assuming that such a relationship *does* exist – and taking into consideration the low recognition rate for *sadness* clips (47.7%) – it is possible to speculate that participants in the DSPD group expressed sadness (facially) to signal their frustration if they were unable to decode the sadness clips. That such a finding was restricted to the DSPD group may say more about differing demand characteristics, rather than indicating a generalisable characteristic of DSPD. Specifically, participants in the DSPD group were not blind to the fact that they formed the experimental group in this research; moreover, these participants were engaged in a pilot treatment programme involving extensive assessments; they may therefore have placed greater personal importance on their performance in the task than either of the control groups, who were similarly un-blinded.

This is not the first study to use the facial affect recognition paradigm with samples possessing elevated PCL-R scores; however, none of the previous studies have employed a format that permits multiple responses on each trial. The finding that the university controls selected more emotional classes per clip than either of the two prison samples is consistent with the hypothesis that individuals with low PCL-R scores are more sensitive to complex emotional expressions, i.e., blends of more than one emotion. Comparing the two prison samples on this measure, the lack of a difference again leaves the uncontrolled variables that may distinguish the university controls from either prison group as potential contributing factors. For example, differences in motor speed may have resulted in participants in the prison samples finding it difficult to enter multiple responses quickly enough before the next trial commenced. However, taking into account the fact that the majority of participants in all three groups labelled at least some clips with more than one emotion, it can be assumed that the time available to make responses was sufficient. This

finding could be pursued in future research, given the increased ecological validity of faces depicting emotional blends over the prototypical faces that have been used previously.

Within- and between-group variability

The findings that came closest to significance were between the university and either one or both of the prison groups, while there were no group differences between the two prison samples. It is likely, based on studies of comparable populations, that the groups differed as predicted in terms of psychopathy scores and severity of personality disorders, i.e., that these were highest in the DSPD group and lowest in the university controls. For example, Forth, Brown, Hart, and Hare (1996) obtained mean PCL: SV score of 6.4 (SD = 5.03) in a sample of 75 male university students, while Cooke, Michie, Hart, and Clark (2005) reported a mean PCL-R score of 16.1 (SD = 8.3) in a sample of 1,316 adult male offenders in the United Kingdom. These figures are substantially lower than the known mean PCL-R score of the DSPD group in this research (28.4; SD = 3.6), and even the lowest DSPD group PCL-R score (24) lies more than one standard deviation above the Cooke et al. (2005) forensic sample. Similarly, the Office for National Statistics reports on psychiatric morbidity (Singleton, Bumpstead, O'Brien, Lee, and Meltzer, 2001; Singleton, Meltzer, and Gatward, 1998) estimated the prevalence for *any* personality disorder is approximately 5% in males living in private households (N = 638) and 64% in male sentenced prisoners (N = 258). These figures are again substantially lower than the known frequency of personality disorder diagnoses in the DSPD group: ten of the eleven participants (91%) had received at least one such diagnosis. It is also likely that the way in which prisoners get onto the DSPD Programme entails that they have been identified by professionals as presenting somewhat differently to the general prison population. However, because the *actual* PCL-R scores and diagnoses were not available for the control groups – and the self-reported figure of only one personality disorder from the prison control group is highly likely to be an underestimate – it is not clear that the groups did in fact differ sufficiently on these variables. Moreover, it is possible that the university sample differed from the two prison samples on a number of variables that were not measured, including level of

education and ethnicity, because of researcher oversight. These additional, unknown sources of variance are a limitation of this research.

A further complicating factor concerns the inherent heterogeneity in a group falling under the working definition of "Dangerous people with severe personality disorder". There may be clinical merits to the terms *dangerous* and *severe* in distinguishing personality disorders with relatively little impact on services from those which are a major concern (Tyrer, 2004); however, the underlying pathologies that characterize this population may be sufficiently diverse to obscure any specific deficits that are pertinent to a more narrowly defined subgroup. In addition to consistently high PCL-R scores, the DSPD group also had a range of personality disorder diagnoses, predominantly within Cluster B but also comorbid with Cluster A or B; two participants had diagnoses from all three clusters. There is evidence for similarities between different personality disorders for certain aspects of emotional processing, such as the physiological responses of patients with BPD and patients with avoidant personality disorder to standardized emotional stimuli (Herpertz et al., 2000) or the reduced facial responsiveness of forensic patients with either BPD or psychopathy when viewing pleasant or unpleasant slides (Herpertz, Dietrich, Wenning, Krings, Erberich, Willmes, Thron, and Saβ, 2001). However, there are also indications that different diagnoses can give rise to contrasting responses in situations with an affective component, such as the finding in the Herpertz et al. (2001) study that, while the BPD and psychopathic samples were similar in their lack of facial responsiveness, patients with BPD (unlike the psychopathic sample) showed electrodermal responses similar to the non-clinical sample. Also, Arntz et al. (2003) found that patients with BPD experience heightened subjective emotional responses to footage depicting abuse relative to patients with Cluster C personality disorders or non-clinical controls. With a phenomenon such as the mirroring of facial expressions of emotion, there may be multiple factors influencing its presence or absence in different individuals who meet the current working definition of DSPD; within-group variability may therefore be more a reflection of pathology than the distribution of an attribute within an homogeneous population.

Strengths and weaknesses of the stimulus set

One of the key strengths of this research lay in the use of dynamic, spontaneous expressions of emotion, in contrast with the still, posed images used in the vast majority of emotion recognition and mirroring studies (e.g., Izard, 1971; Ekman and Friesen, 1971; Kosson et al., 2002; Wallbott, 1991, Sonnby-Borgström, 2002; Sonnby-Borgström et al., 2003). As there are currently no widely available, validated stimulus sets of this nature – and creating such a set was beyond the scope of this research – clips were selected from an existing database. Despite this containing a large number of video clips depicting different expressions of emotion, it was not possible to equate perceived strength ratings across emotional classes, without the resulting stimulus set comprising only weak examples of each emotional class. The findings of (close to significant) between-group differences in mirroring and recognition for specific emotions are confounded by this fact; however, as the group differences in mirroring were associated with the strongest clips, and the differences in recognition were associated with the weakest clips, it is possible to speculate that the differences observed may be more global.

There remains a strong case for using dynamic facial expressions of emotion: they have far more in common with the typical faces that people encounter in everyday life than do photographs. However, capturing a large enough number of spontaneous expressions in a controlled environment – that are sufficiently extreme to induce mirroring – may be impractical. An alternative approach, constituting something of a compromise, would be to compile a database of dynamic, *posed* expressions. Individuals with some formal training in acting would potentially make suitable participants, given that one of the core attributes of an effective actor is an ability to pass off as genuine what are in fact artificial emotional expressions.

Measurement of mirroring

Mirroring was low in all three groups and across emotional classes. While this is likely to reflect in part the weakness of some of the clips and the naturally low prevalence of the phenomenon, it may also be a consequence of the method used to measure mirroring. EMG would have provided the most accurate measurements of muscular movements, and is the method recommended by Tassinary and Cacioppo (1992) to detect weak facial expressions. However, one of the aims of this study was to develop a means to measure mirroring that could be incorporated into clinical assessment and treatment evaluation. EMG requires both specialist equipment and training that are prohibitively expensive. The Facial Action Coding System (FACS) provides a highly structured procedure for measuring facial expressions without the need for equipment, but again specialist training is needed. A further, conceptual argument can be made against these alternative methods: assuming that mirroring has a social, communicative purpose, it is those responses that are detectable by untrained (but socialized) humans that truly constitute the phenomenon of interest, rather than imperceptible muscular twitches or complex displays that need to be deciphered using a manual.

The sound theoretical rationale for the method employed here notwithstanding, the low interrater reliability implies that some form of coaching would have benefitted raters. It may have sufficed to have shown them a sample of prototypical faces prior to carrying out the rating, or alternatively to have asked raters to attempt the same facial affect recognition task as the participants completed. This latter suggestion could have provided an effective way to sensitize the raters to the subtle echoes displayed on the faces of the participants.

In addition to the two sets of mirroring scores generated by having two raters, scores were also reported in both a *raw* and a *composite* form. The raw scores indicated the percentage of trials on which there was a match between the emotional class of the clip being viewed, and the label selected by the rater. The composite scores took the raw measures for any particular emotional class, and subtracted the percentage of clips corresponding to *neutral* trials that were labelled by raters as the same particular class. This adjustment was made in

order to control for possible group differences in natural facial expressions that had nothing to do with responding to particular classes of stimulus. Only one significant group difference disappeared as a consequence of the adjustment: the increased tendency of participants in the DSPD group to mirror *anger*. It could be argued that changing the scores in this way in fact eliminated a real effect, i.e., that individuals with DSPD do in fact mirror *anger* in an exaggerated way. Theoretically, such an effect could be a consequence of elevated trait anger in this group, although as this was not measured such an interpretation is purely speculative. Moreover, the effect was only present in the data from Rater 1 who, as was stated earlier, was not perfectly blind to the identity of members of the DSPD group. It may therefore be an artifact of unintentional bias.

Implications

In order for readily observable mirroring to occur, stimuli are likely to have to comprise strong examples of facial expressions of emotion, which may only be practicably obtainable by resorting to posed expressions. Power calculations for the effect sizes reported here (*Range* of $r = 0.31 - 0.39$) indicate that sample sizes of 88 – 138 would be required to achieve a statistical power of 0.95, which although not large compared with many studies, may be impractical at present given the small DSPD population in England, which runs into the hundreds rather than thousands. Moreover, exposing a severely personality disordered sample of this size to even extreme and exaggerated expressions may not result in mirroring, a finding that would be consistent with research investigating the association between mirroring and empathy (Sonnby-Borgström, 2002; Sonnby-Borgström et al., 2003). Whether the mechanism (or mechanisms) underlying this and other affective deficits seen in DSPD would be amenable to therapeutic intervention is at present unknown. There is a growing body of research suggesting that personality is not, as was once thought, stable over long periods of time (e.g., Shea et al., 2002; Johnson, Cohen, Kasen, Skodol, Hamagami, and Brook, 2000; Seivewright et al., 2002). Treatment that aims to rehabilitate the interpersonal and affective functioning of people currently defined as dangerous and having a severe personality disorder, may result in the normalisation of such

behaviours as the mirroring of facial expressions of emotion. Alternatively, simply informing individuals that they have a deficit in mirroring, i.e., bringing it into their conscious awareness, may itself constitute a direct way to target this difference in one aspect of social functioning. The procedure described here, incorporating the refinements suggested, could provide a means to assess mirroring. If it was possible to overcome the limitations of sample size, it could also be used in a dismantling study, with pre- and post-treatment testing being applied to compare groups receiving regular treatment and those receiving treatments targeted specifically at improving mirroring.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and Statistical Manual: 4th Edition Text Revision*. Washington, DC: American Psychiatric association.

Barnard, P.J. & Teasdale, J.D. (1991). Interacting cognitive subsystems: A systemic approach to cognitive-affective interaction and change. *Cognition and Emotion*, *5*, 1-39.

Blair, R.J.R., (2001). Neuro-cognitive models of aggression, the antisocial personality disorders and psychopathy. *Journal of Neurology, Neurosurgery and Psychiatry, 71,* 727–731.

Blair, R. J. R., Colledge, E., Murray, L., & Mitchell, D. G. V. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology, 29,* 491-498.

Blair, R. J. R., Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology, 34*, 192-198.

Blair, R.J.R., Mitchell, D.G.V., Peschardt, K.S., Colledge, E., Leonard, R.A., Shine, J.H., Murray, L.K., & Perrett, D.I., 2004. Reduced sensitivity to other's fearful expressions in psychopathic individuals. *Personality and Individual Differences, 37*, 1111 – 1122.

Cacioppo, J. T., Petty, R. E., Losch, M., & Kim, H. S. (1986), Electromyographic activity over facial muscle regions can differentiate the valence and intensity of emotional reactions. *Journal of Personality and Social Psychology, 50,* 260-268.

Chemtob, C. M., Novaco, R. W., Hamada, R. S., & Gross, D. M. (1997). Cognitive-behavioural treatment for severe anger in posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 65*, 184–189.

Cleckley, H. *The Mask of Sanity. 5th Edn.* St Louis, MO: Mosby.

Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, 53, 751-766.

Cohen, D. & Strayer, J. (1996). Empathy in conduct-disordered and comparison youth. *Developmental Psychology, 32,* 988-998.

Cooke, D. J., & Michie, C. (1997). An item response theory analysis of the Hare Psychopathy Checklist-Revised. *Psychological Assessment, 9*, 3-14.

Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopath: Towards a hierarchical model. Psychological Assessment, 13, 171-188.

Cooke, D. J., Michie, C., Hart, S. D., & Clark, D. (2005). Assessing Psychopathy in the UK: Concerns about cross-cultural generalisability. *British Journal of Psychiatry, 186,* 335-341.

Darwin, C. (1872). *The Expression of Emotions in Man and Animals*. New York: D. Appleton and Company.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. JSAS: *Catalog of Selected Documents in Psychology, 10*, 85. (Ms. No. 2124).

Department of Health, Home Office. (1999). *Managing Dangerous People with Severe Personality Disorder.* London: Home Office.

Dimberg, U., 1982. Facial reactions to facial expressions. *Psychophysiology, 19*, 643-647.

Dimberg, U., 1997. Facial reactions: rapidly evoked emotional responses. *Journal of Psychophysiology, 11,* 115-123.

Dolan, B., & Coid, J. (1993). *Psychopathic and Antisocial Personality Disorders: Treatment and Research Issues.* London: Gaskell.

Duan, C., & Hill, C. E. (1996). The current state of empathy research. *Journal of Counseling Psychology, 43*, 261-274.

Eisenberg, N., Fabes, R., Bustamante, D., & Mathy, R. (1987). Physiological indices of empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its Development*. New York: Cambridge University Press.

Ekman, P. (1999). Facial expressions. In T. Dalgleish & M. Power (Eds.) *Handbook of Cognition and Emotion*. New York: Wiley.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124-129.

Ekman, P, & Friesen, W. V. (1976). *Pictures of Facial Affect*. Palo Alto, California: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. New York: Pergamon Press.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 221,* 1208-1210.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, California: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K. R., Tomita, M., &

Tzavaras, A. (1987). Universals and cultural differences in the judgements of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*, 712-717.

Feschbach, N. D., & Roe, K. (1968). Empathy in six- and seven-year-olds. *Child Development, 39*, 133-145.

Field, A. (2005). *Discovering Statistics Using SPSS (Second Edition)*. London: SAGE.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23*, 567-589.

Fulero, S. M. (1995). Review of the Hare Psychopathy Checklist-Revised. In J. C. Conoley & J. C. Impara (Eds.), *Twelfth Mental Measurements Yearbook* (453-454). Lincoln, NE: Buros Institute.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behaviour*, *29*, 397-426.

Goodenough, W. H. (1956). Componential analysis and the study of meaning. *Language, 32,* 195-216.

Graham, J. L. (1980). A new system for measuring nonverbal responses to marketing appeals. *1980 AMA Educator's Conference Proceedings, 46,* 340-343.

Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition and Emotion, 13,* 225–266.

Hare, R. D. (1991). *The Hare Psychopathy Checklist – Revised*. Toronto: Multi-Health Systems.

Hare, R. D. (2003). *Hare Psychopathy Checklist – Revised (PCL-R), 2nd Ed. Technical manual*. New York: Multi-Health Systems.

Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P., (1990). The Revised Psychopathy Checklist: Descriptive statistics, reliability, and factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2, 338-341.

Hare, R. D., Hart, S. D., & Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *Journal of Abnormal Psychology, 100,* 391-398.

Harris, G. T., Rice, M. E., & Cormier, C. A. (1991). Psychopathy and violent recidivism. *Law and Human Behavior, 15*, 625–637.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20*, 315-335.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1994). Psychopathy as a taxon: Evidence that psychopaths are a discrete class. *Journal of Consulting and Clinical Psychology*, *62*, 387-397.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *Manual for the Hare Psychopathy Checklist: Screening Version (PCL: SV)*. Toronto, Canada: Multi-Health Systems.

Hemphill, J. F., & Hare, R. D. (2004). Some misconceptions about the Hare PCL-R and risk assessment: A reply to Gendreau, Goggin, and Smith. *Criminal Justice and Behavior, 31*, 203-243.

Herpertz, S. C., Schwenger, U. B., Kunert, H. J., Lukas, G., Gretzer, U., Nutzmann, J., Schuerkens, A., & Sass, H. (2000). Emotional responses in patients with borderline as compared with avoidant personality disorder. *Journal of Personality Disorders, 14*, 339-351.

Herpertz, S. C., Werth, U., Lukas, G., Qunaibi, M., Schuerkens, A., Kunert, H. J., Freese, R., Flesch, M., Mueller-Isberner, R., Osterheider, M., & Sass, H. (2001). Emotion in criminal offenders with psychopathy and borderline personality disorder. *Archives of General Psychiatry, 58*, 737-745.

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology, 40*, 129-141.

Hess, U., Blairy, S., Philippot, P. (1999). Facial mimicry. In: Phillippot, P., Feldman, R., Coats, E. (Eds.), *The Social Context of Nonverbal Behaviour*. New York: Cambridge University Press.

Hoffman, M. L. (1984). Interaction of affect and cognition on empathy. In: Izard, C. E., Kagan, J., Zajonc, R. B. (Eds.), *Emotion, Cognition, and Behaviour*. Cambridge: Cambridge University Press.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33*, 307-316.

Howell, D. C. (1999). *Fundamental statistics for the Behavioral Sciences: 4th Edn.* California: Brooks/Cole Publishing Company.

Izard, C. (1971). *The Face of Emotion.* New York: Appleton-Century-Crofts.

Johnson, J G., Cohen, P., Kasen, S., Skodol, A. E., Hamagami, E., & Brook, J. S. (2000). Age-related change in personality disorder trait levels between early adolescence and adulthood: A community-based longitudinal investigation. *Acta Psychiatrica Scandinavica, 102,* 265-275.

Kosson, D. S., Suchy, Y., Mayer, A. R., & Libby, J. (2002). Facial affect recognition in criminal psychopaths. *Emotion, 2*, 398-411.

Kraemer, H. C. (1979). Ramifications of a population model for κ as a coefficient of reliability. *Psychometrika, 44,* 461-472.

Lanzetta, T., & Englis, B. G. (1989). Expectations of cooperation and competition and their effects on observers' vicarious emotional responses. *Journal of Personality and Social Psychology, 56*, 543-554.

Leventhal, H. (1984). A perceptual-motor theory of emotion. *Advances in Experimental Social Psychology, 17,* 117-182.

Lindman, H. R. (1974). *Analysis of Variance in Complex Experimental Designs*. San Fransisco: Freeman.

Lipps, T. (1903) Einfühlung, innere Nachahmung und Organempfindung. *Archiv für die gesamte Psychologie, 1,* 465-519. Cited in Wispe, L. (1987). History of the concept of empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its Development*. New York: Cambridge University Press.

Lipps, T. (1907). Das Wissen von fremden Ichen. In: Lipps, T. (Ed.), Psychologische Untersuchungen (Band 1). Engelmann, Leipzig. Cited in Hess, U., & Blairy, S. (2002). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology, 40*, 129-141.

Listowel, E. (1934). *A Critical History of Aesthetics*. London: G. Allen and Unwin.

Livesley, J. W. (2001). *Handbook of Personality Disorders*. New York: Guilford Press.

Lundqvist, L. O., & Dimberg U. (1995). Facial expressions are contagious. *Journal of Psychophysiology, 9,* 203-211.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*, 525-543.

Mikhailova, E.S., Vladimirova, T.V., Iznak, A.F., Tsusulkovskaya, E.J., & Sushko, N.V., 1996. Abnormal recognition of facial expression of emotions in depressed patients with major depression disorder and schizotypal personality disorder. *Biological Psychiatry, 40,* 697-705.

Miller, P. A., & Eisenberg, N. (1988). The relation of empathy to aggressive and externalizing/antisocial behaviour. *Psychological Bulletin, 103,* 324-344.

Moltó, J., Poy, R., & Torrubia, R. (2000). Standardization of the Hare Psychopathy Checklist-Revised in a Spanish prison sample. *Journal of Personality Disorders, 14*, 84-96.

Ogloff, J. R., Wong, S., & Greenwood, A. (1990). Treating adult psychopaths in a therapeutic community program within a correctional setting. In Cooke, D. J., Michie, C., Hart, S. D., & Hare, R. D. (1999). Evaluating the screening version of the Hare Psychopathy Checklist – Revised (PCL: SV): An item response theory analysis. *Psychological Assessment*, 11 (1), 3-13.

Oldham, J. M., Skodol, A. E., Kellman, H. D., Hyler, S. E., Rosnick, L., Davies, M. (1992). Diagnosis of DSM-III-R personality disorders by two structured interviews: Patterns of comorbidity. *American Journal of Psychiatry, 149,* 213-220.

O'Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H., & Abdi, H. (2005). A Video Database of Moving Faces and People. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27,* 812-816.

Ortony, A., Clore, G. L., & Foss, M. A. (1987). The referential structure of the affective lexicon. *Cognitive Science, 11*, 341-364.

Power, M. J., & Dalgleish, T. (1997). *Cognition and Emotion: From Order to Disorder.* Hove: Psychology Press.

Renneberg, B., Heyn, K., Gebhard, R., & Bachmann, S. (2005). Facial expression of emotions in borderline personality disorder and depression. *Journal of Behavior Therapy and Experimental Psychiatry, 36,* 183-196.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95-103.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of cross-cultural studies. *Psychological Bulletin, 115*, 102-141.

Russell, J. A. (1995). Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin, 118*, 379-399.

Russell, J. A., & Yik, M. S. M. (1996). Emotion among the Chinese. In M. H. Bond (Ed.), *The Handbook of Chinese Psychology*. Hong Kong: Oxford
University Press.

Salekin, R., Rogers, R., & Sewell, K. (1996). A review and meta-analysis of the psychopathy checklist and psychopathy checklist-revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice, 3*, 203-215.

Seivewright, H., Tyrer, P., & Johnson, T. (2002). Change in personality status in neurotic disorders. *Lancet, 359,* 2253-2254.

Shea, M. T., Stout, R. L., Gunderson, J., Morey, L., Grilo, C. M., McGlashan, T. H., Skodol, A. E., Dolan-Sewell, R., Dyck, I., Zanarini, M. C., & Keller, M. B. (2002). Short-term diagnostic stability of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *American Journal of Psychiatry, 159,* 2036–2041.

Shea, M. T., & Yen, S. (2003). Stability as a distinction between Axis I and Axis II disorders. *Journal of Personality Disorders*, 17, 373-386.

Singleton, N., Bumpstead, R., O'Brien, M., Lee, A. & Meltzer, H. (2001). *Psychiatric Morbidity Among Adults Living in Private Households, 2000*. London: Office of National Statistics.

Singleton, N., Meltzer, H. & Gatward, R. (1998) *Psychiatric Morbidity Among Prisoners in England and Wales*. London: Office for National Statistics.

Smith, C. A., McHugo, G. J., & Lanzetta, J. T. (1986). The facial muscle patterning of posed and imagery-induced expressions of emotion by expressive and nonexpressive posers. *Motivation and Emotion, 10,* 133-157.

Sonnby-Borgström, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scandinavian Journal of Psychology, 43,* 433-443.

Sonnby-Borgström, M., Jönsson, P., & Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behaviour, 27,* 3-23.

Stone, G. L. (1995). Review of the Hare Psychopathy Checklist-Revised. In J. C. Conoley & J. C. Impara (Eds.), *Twelfth Mental Measurements Yearbook* (454-455). Lincoln, NE: Buros Institute.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics (4th Edn).* Needham Heights, MA: Allyn and Bacon.

Tassinary, L. G., & Cacioppo, J. T. (1992). Unobservable facial actions and emotion. *Psychological Science, 3,* 28-33.

Titchener, E. (1924). *A Textbook of Psychology*. New York: Macmillan.

Truax, C. B., & Carkhuff, R. R. (1967). *Toward Effective Counseling and Psychotherapy: Training and Practice*. Chicago: Aldine.

Tyrer, P. (2004). Getting to grips with severe personality disorder. *Criminal Behaviour and Mental Health, 14*, 1-4.

Van IJzendoorn, M. H., Feldbrugge, J. T. T. M., Derks, F. C. H., De Ruiter, C., Verhagen, M. F. M., Philipse, M. W. G., Van der Staak, C. P. F., & Riksen-Walraven, J. M. A. (1997). Attachment representations of personality-disordered criminal offenders. *American Journal of Orthopsychiatry, 67,* 449-459.

Vitale, J. E., & Newman, J. P. (2001). Using the Psychopathy Checklist-Revised with female samples: Reliability, validity, and implications for clinical utility. *Clinical Psychology: Science, & Practice,* 8, 117-132.

Wallbott, H.G., 1991. Recognition of emotion from facial expression via imitation? Some indirect evidence for an old theory. *British Journal of Social Psychology, 30*, 207_219.

Warren, F., Preedy-Fayers, K., McGauley, G., Pickering, A., Norton, K., Geddes, J., & Dolan, B. (2003). *Review of Treatments for Severe Personality Disorder*. London: Home Office.

Watt, B. D., & Howells, K. (1999). Skills training for aggression control: evaluation of an anger management programme for violent offenders. *Legal and Criminological Psychology, 4,* 285–300.

Wixon, D., & Laird, J. D. (1981). Individual differences and the effects of mimicry on the judgement of facial expressions. Unpublished manuscript cited in Lanzetta, J. T., & Englis, B. G. (1989). Expectations of cooperation and competition and their effects on observers' vicarious emotional responses. *Journal of Personality and Social Psychology, 56*, 543-554.

Zagon, I. K., & Jackson, H. J. (1994). Construct validity of a psychopathy measure. *Personality and Individual Differences, 17,* 125-135.

APPENDIX A: Ethical approval

Telephone enquiries, please contact
Cynthia Richardson on 01924 213110
Fax 01924 213195
e-mail Cynthia.Richardson@wwpct.nhs.uk

Wakefield West
Primary Care Trust

White Rose House
West Parade
Wakefield
WF1 1LT

Revised Version

27th September 2005

Mr Keith W Whittle
Psychologist in Clinical Training
University of Leeds
Academic Unit of Psychiatry & Behavioural Sciences
15 Hyde Terrace
Leeds
LS2 9LT

Dear Mr Whittle

Full title of study: **Facial responses to facial expressions of emotion of prisoners with dangerous and severe personality disorder**
REC reference number: **05/Q1204/42**

Thank you for your letter of the 5th September 2005, responding to the Committee's request for further information on the above research and the revised documentation.

The further information has been considered on behalf of the Committee by the Vice-Chairman, Dr Susan Macdonald-Hull.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

05/Q1204/42

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Application	1	22 July 2005
Investigator CV	1	22 July 2005
Protocol	1	22 July 2005
Covering Letter	1	22 July 2005
Covering Letter		05 September 2005
Letter from Sponsor Cambridgeshire & Peterborough		26 August 2005
Peer Review Letter from Natercia Godinho		26 August 2005
Participant Information Sheet	2	05 September 2005
Participant Information Sheet Prisoner Control Group	2	05 September 2005
Participant Information Sheet	2	05 September 2005
Participant Information Sheet Prisoner Control Group	2	05 September 2005
Participant Information Sheet	1	22 July 2005
Participant Information Sheet (Prisoner Control Group)	1	22 July 2005
Participant Information Sheet Control Group	2	05 September 2005
Participant Consent Form	1	22 July 2005
Participant Consent Form University participants	1	05 September 2005
Participant Consent Form	2	05 September 2005
Participant Consent Form University Participants	1	05 September 2005
Participant Consent Form	2	05 September 2005
Letter of Information	1	22 July 2005
Checklist		

Research governance approval

The study should not commence at any NHS site until the local Principal Investigator has obtained final research governance approval from the R&D Department for the relevant NHS care organisation.

You should arrange for all relevant NHS care organisations to be notified that the research will be taking place, and provide a copy of the REC application, the protocol and this letter.

All researchers and research collaborators who will be participating in the research must obtain research governance approval from the relevant care organisation before commencing any research procedures. Where a substantive contract is not held with the care organisation, it may be necessary for an honorary contract to be issued before approval for the research can be given.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

05/Q1204/42

With the Committee's best wishes for the success of this project

Yours sincerely

C M Richardson

Cynthia M Richardson (Mrs)
Committee co-ordinator

Enclosures *Standard approval conditions*

Copy to *R&D Department, University of Leeds*

APPENDIX B: Letter of invitation (DSPD)

Date

Dear _____,

I am a Psychologist in Clinical Training at the University of Leeds, and I am currently carrying out some research at HMP Whitemoor. I am investigating how prisoners on the DSPD Programme respond to facial expressions of emotion, such as "happiness", "anger", "surprise", etc. I wish to work with people with a diagnosis of personality disorder when my training is completed, and I am keen to be involved in a project that can help improve understanding of what it means to have such a diagnosis.

Details of the proposed project can found on the enclosed Participant Information Sheet. Should you wish to take part, it is estimated that your participation will take around half an hour. Your participation is voluntary and not participating will not affect your normal care in any way. Also, you are free to end your participation at any time and are not expected to give any reason for why you have changed your mind.

I hope that you will consider taking part in this research project, as in addition to increasing our understanding of what it means to have a diagnosis of severe personality disorder, it may also aid in the evaluation of the treatment programme with which you are currently involved.

If you have further questions about the research that you wish to be answered, you are welcome to ask **Dr Naomi Murphy**, who has agreed to act as my point of contact on the wing. However, all prisoners will be given the opportunity to have any questions answered directly by myself before agreeing to take part.

Yours sincerely,

Keith Whittle
Psychologist in Clinical Training
University of Leeds

APPENDIX C: Letter of invitation to (prison control)

Date

Dear _____,

I am a Psychologist in Clinical Training at the University of Leeds, and I am currently carrying out some research at HMP Whitemoor. I am investigating how prisoners on the Dangerous and Severe Personality Disorder (DSPD) Programme respond to facial expressions of emotion, such as "happiness", "anger", "surprise", etc. I am writing to you, as a person who does *not* have a diagnosis of severe personality disorder, to invite you to form part of a group that would be compared with a group of DSPD prisoners and also with a group of people who are not prisoners. In this way, I would be able to see how the responses of people with a diagnosis of severe personality disorder might differ from people such as yourself who do not have such a diagnosis. Your responses would not be used to assess you personally in any way.

Details of the proposed project can found on the enclosed Participant Information Sheet. Should you wish to take part, it is estimated that your participation will take around half an hour. Your participation is voluntary and not participating will not affect your normal care in any way. Also, you are free to end your participation at any time and are not expected to give any reason for why you have changed your mind.

I hope that you will consider taking part in this research project, as it may make a valuable contribution to our understanding of what it means to have a diagnosis of severe personality disorder.

If you have further questions about the research that you wish to be answered, you are welcome to ask SO Dave Watson or SO Barry Leban, who have agreed to act as my points of contact on the wing. However, all prisoners will be given the opportunity to have any questions answered directly by myself before agreeing to take part.

Yours sincerely,

Keith Whittle
Psychologist in Clinical Training
University of Leeds

APPENDIX D: Participant information sheet (DSPD)

PARTICIPANT INFORMATION SHEET

RESPONSES TO FACIAL EXPRESSIONS OF EMOTION

Information about the project:

I am a researcher in the Institute of Psychological Sciences at the University of Leeds. I am interested in how people with a diagnosis of Dangerous and Severe Personality Disorder (DSPD) recognise and react to emotions in others.

This research is subject to ethical guidelines set out by the British Psychological Society. These include obtaining your informed consent, making clear that you are free to withdraw your consent at any time before the research is completed, and ensuring that you cannot be identified by name in the final results. This sheet will hopefully provide you with enough information about the study to allow you to make an informed decision about participation. However, if you have any questions or would like to discuss anything with me please let me know.

What it will involve

I would like you to take part because I am approaching you as a potential participant, as you are currently involved in the DSPD treatment programme at HMP Whitemoor. I am interested in looking at your responses to a task in which you are asked to watch a series of short video clips of human faces. During the task, you would be asked to label the emotion that you feel is being expressed by each face, such as "happiness", "anger", "surprise", etc. You would be filmed during the task, and the recording would be used to help me to analyse your responses. The recording would only be viewed by the following people: myself, my research supervisors, and a small group of prison staff. Your responses will then be compared with those of a group of prisoners at HMP Whitemoor who are not involved in the DSPD programme, and also with a group of people from the community.

What will happen to the information?

The information about your responses will be made anonymous and combined with the information from other participants, so that it will not be possible to name or identify you at all in any final reports. These reports could include presentations at conferences and articles in scientific journals.

The video footage will be destroyed no more than 12 months after the study is completed, and will not be used as part of any presentation.

Benefits of participating

Participating in this research is voluntary and the choice not to participate will not affect your normal care or treatment in any way.

It is hoped that the findings of this research will help to improve understanding about how people with DSPD experience and make sense of emotions. The task that you are being asked to do may also be developed to help evaluate the progress that prisoners make in treatment at HMP Whitemoor.

I would be very grateful for any help you can give me with this research. If you wish to withdraw your consent to participate at any time, you are free to do so and your decision will be respected without you having to give a reason, and without affecting your normal care or treatment in any way.

Further questions

Should you have any further questions about this research, I will be happy to meet with you and discuss them before you agree to take part.

APPENDIX E: Participant information sheet (prison control)

PARTICIPANT INFORMATION SHEET (PRISONER CONTROL GROUP)

RESPONSES TO FACIAL EXPRESSIONS OF EMOTION

Information about the project:

I am a researcher in the Institute of Psychological Sciences at the University of Leeds. I am interested in how people with a diagnosis of Dangerous and Severe Personality Disorder (DSPD) recognise and react to emotions in others.

This research is subject to ethical guidelines set out by the British Psychological Society. These include obtaining your informed consent, making clear that you are free to withdraw your consent at any time before the research is completed, and ensuring that you cannot be identified by name in the final results. This sheet will hopefully provide you with enough information about the study to allow you to make an informed decision about participation. However, if you have any questions or would like to discuss anything with me please let me know.

What it will involve

I would like you to take part because I am approaching you as a potential research participant to help form part of a group that will be compared with a group of prisoners on the DSPD programme at HMP Whitemoor, and also with another group of people from the community. The contribution from you would be combined with that of other people in your group, so that you would not be personally identifiable in any final reports. These combined results would then be used to help me make sense of the results from the other two groups. Your responses would not be used to assess you personally in any way.

I am interested in looking at your responses to a task in which you are asked to watch a series of short video clips of human faces. During the task, you would be asked to label the emotion that you feel is being expressed by each face, such as "happiness", "anger", "surprise", etc. You would be filmed during the task, and the recording would be used to help me to analyse your responses. The recording would only be viewed by the following people: myself, my research supervisors, and a small group of prison staff.

What will happen to the information?

The information about your responses will be made anonymous and combined with the information from other participants, so that it will not be possible to name or identify you at all in any final reports. These reports could include presentations at conferences and articles in scientific journals.

The video footage will be destroyed no more than 12 months after the study is completed, and will not be used as part of any presentation.

Benefits of participating

Participating in this research is voluntary and the choice not to participate will not affect your normal care or treatment in any way.

It is hoped that the findings of this research will help to improve understanding about how people with DSPD experience and make sense of emotions. The task that you are being asked to do may also be developed to help evaluate the progress that prisoners in the DSPD Programme make in treatment at HMP Whitemoor.

I would be very grateful for any help you can give me with this research. If you wish to withdraw your consent to participate at any time, you are free to do so and your decision will be respected without you having to give a reason, and without affecting your normal care or treatment in any way.

Further questions

Should you have any further questions about this research, I will be happy to meet with you and discuss them before you agree to take part.

APPENDIX F: Participant information sheet (university control)

PARTICIPANT INFORMATION SHEET (CONTROL GROUP)

RESPONSES TO FACIAL EXPRESSIONS OF EMOTION

Information about the project:

I am a researcher in the Institute of Psychological Sciences at the University of Leeds. I am interested in how people with a diagnosis of Dangerous and Severe Personality Disorder (DSPD) recognise and react to emotions in others.

This research is subject to ethical guidelines set out by the British Psychological Society. These include obtaining your informed consent, making clear that you are free to withdraw your consent at any time before the research is completed, and ensuring that you cannot be identified by name in the final results. This sheet will hopefully provide you with enough information about the study to allow you to make an informed decision about participation. However, if you have any questions or would like to discuss anything with me please let me know.

What it will involve

I would like you to take part because I am approaching you as a potential research participant to help form part of a group that will be compared with a group of prisoners on the DSPD programme at HMP Whitemoor, and also with another group of prisoners at HMP Whitemoor who are not on the DSPD programme. The contribution from you would be combined with that of other people in your group, so that you would not be personally identifiable in any final reports. These combined results would then be used to help me make sense of the results from the other two groups. Your responses would not be used to assess you personally in any way.

I am interested in looking at your responses to a task in which you are asked to watch a series of short video clips of human faces. During the task, you would be asked to label the emotion that you feel is being expressed by each face, such as "happiness", "anger", "surprise", etc. You would be filmed during the task, and the recording would be used to help me to analyse your responses. The recording would only be viewed by the following people: myself, my research supervisors, and a small group of prison staff.

What will happen to the information?

The information about your responses will be made anonymous and combined with the information from other participants, so that it will not be possible to name or identify you at

all in any final reports. These reports could include presentations at conferences and articles in scientific journals.

The video footage will be destroyed no more than 12 months after the study is completed, and will not be used as part of any presentation.

Benefits of participating

It is hoped that the findings of this research will help to improve understanding about how people with DSPD experience and make sense of emotions. The task that you are being asked to do may also be developed to help evaluate the progress that prisoners in the DSPD Programme make in treatment at HMP Whitemoor.

I would be very grateful for any help you can give me with this research. If you wish to withdraw your consent to participate at any time, you are free to do so and your decision will be respected without you having to give a reason, and without affecting your legal rights in any way.

Further questions

Should you have any further questions about this research, I will be happy to answer them before you agree to take part.

APPENDIX G: Informed consent form (prison)

INFORMED CONSENT FORM

FACIAL EXPRESSIONS OF EMOTION INVESTIGATION

The purpose of this form is to help us ensure you are willing to take part in this study and understand our willingness to accommodate you in any way we can. Because we need to keep a record of this the language is rather more formal than we would like. Signing this form does not commit you to anything you do not wish to do.

Please circle as appropriate

- Have you read the participant information sheet? Yes / No

- Have you had the opportunity to ask questions and discuss the study? Yes / No

- Have you received satisfactory answers to your questions? Yes / No

- Do you understand that your participation is voluntary and you are free to withdraw:
  - At any time? Yes / No
  - Without having to give a reason for withdrawing? Yes / No
  - Without having your care, treatment or legal rights affected in any way? Yes / No

- Do you agree to be video-recorded during this study? Yes / No
(The recordings will be viewed only by the researchers and a panel of raters, who will be staff at the prison. All video recordings will be destroyed after completion of the study).

- Do you agree to take part in this study? Yes/No

Signed…… **Date**……

Name in block capitals:……

APPENDIX H: Informed consent form (university)

INFORMED CONSENT FORM (UNIVERSITY PARTICIPANTS)

FACIAL EXPRESSIONS OF EMOTION INVESTIGATION

The purpose of this form is to help us ensure you are willing to take part in this study and understand our willingness to accommodate you in any way we can. Because we need to keep a record of this the language is rather more formal than we would like. Signing this form does not commit you to anything you do not wish to do.

Please circle as appropriate

- Have you read the participant information sheet? Yes / No

- Have you had the opportunity to ask questions and discuss the study? Yes / No

- Have you received satisfactory answers to your questions? Yes / No

- Do you understand that your participation is voluntary and you are free to withdraw:
  - At any time? Yes / No
  - Without having to give a reason for withdrawing? Yes / No
  - Without having your legal rights affected? Yes / No

- Do you agree to be video-recorded during this study? Yes / No
(The recordings will be viewed only by the researchers and a panel of raters, who will be staff at HMP Whitemoor. All video recordings will be destroyed after completion of the study).

- Do you agree to take part in this study? Yes/No

Signed..... **Date**.....

Name in block capitals:.....

APPENDIX I: Instructions for the initial rating task

You will be shown a series of short video clips of peoples' faces.

After you have viewed each clip, decide which emotion label best fits the facial expression you have just seen from the following selection:

anger	**disgust**	**fear**	**happiness**	**sadness**	**surprise**	**neutral**

Then rate the strength of the facial expression using the following scale:

1	**2**	**3**	**4**	**5**	**6**	**7**

weak ⟶ **strong**

If you decide that there is no facial expression of emotion, simply state **neutral** and you don't need to then rate the strength.

If you decide that the clip would best be described by more than one of the above terms, state **blend** then indicate which of the above emotions best describes the expression and rate each one on the 7-point scale.

If you decide that the facial expression depicts a different emotion not included in the selection above, state the word that best describes the emotion and rate it on the 7-point scale.

APPENDIX J: Instructions for the main task

- You will be presented with a number of short video clips of peoples' faces.
- After you have seen each face, you will have a short time to choose the emotion word or words that best describe the face that you have just seen.

- The choices are:

Anger Disgust Fear Happiness Sadness Surprise Other Neutral

- Mark your choice on the answer sheet provided by circling a number underneath the emotion word from 1 to 7, where 1 is the weakest and 7 is the strongest expression of that particular emotion.
- For faces that you decide are "neutral", that is, faces that are not expressing an emotion, just circle the word "neutral" on the answer sheet.

- You may decide that more than one emotion word is needed to describe the face.
- If so, just circle a number under each emotion word that is needed from 1 (weak) to 7 (strong).

- You may also decide that a different word would be better to describe the face.
- If so, write this word on the answer sheet where it says "Other (please state)", and circle a number underneath it from 1 (weak) to 7 (strong).

PRACTICE

- To help you get used to the task there will first be a short practice.

- During the practice you will be given more time between each video clip at first, but then the time will get shorter as you get used to the task.

- After the practice, you will be able to do the main task as soon as you are ready.
- You will also be able to have a break halfway through the main task.

- Try not to think about it too much: it's your first answer that we are interested in.

Remember, there are no right answers!

APPENDIX K: Instructions for the final rating task

You will be shown a series of short video clips of peoples' faces.

The people were filmed while also viewing video clips, depicting a range of different facial expressions of emotion that included *anger*, *disgust*, *fear*, *happiness*, *sadness*, *surprise*, and *neutral*.

Sometimes people mirror the facial expressions that they see.

It is your task with each clip to try and guess which emotion the person was seeing. Please do this by choosing **one** of the following labels for each clip:

anger	**disgust**	**fear**	**happiness**	**sadness**	**surprise**	**neutral**

If you cannot see any emotional expression on a face that you are viewing, label it as *neutral*.

Indicate your answers on the response sheet by drawing a line through the label you have chosen for each clip.

APPENDIX L: Transformation of *other* responses

Response (*lexicon entry if different*)	Rational	Empirical	Relevant thesaurus entry(s)	Reclassification
annoyed	affective	affective	annoyed = angry	anger
disapproval (*disapprove-of*)	affective	affective	disapproval = anger	anger
fed up	affective	affective	fed up = annoyed = angry	anger
violent*	cognitive	affective	-	anger
contempt	affective	affective	contempt = disgust	disgust
concern	affective	affective	concern = fear	fear
concerned	affective	affective	concerned = fearful	fear
excited	affective	affective	excited = agitated = worry = fear *or* excited = nervous = fearful	fear
worried	affective	affective	worried = fearful	fear
worry	affective	affective	worry = fear	fear
amused	affective	affective	amused = pleased = happy	happiness
amusement	affective	cognitive	amusement = delight = happiness	happiness
content (*contented*)	affective	affective	content = happy	happiness
mildly amused	affective	affective	amused = pleased = happy	happiness
pleased	affective	affective	pleased = happy	happiness
pride	affective	affective	pride = happiness	happiness
relaxed	affective	affective	relaxed = happy	happiness
serene	affective	external	serene = contented = happy	happiness
smugness (*smug*)	affective	cognitive	smug = pleased = happy	happiness

* This word occurred on only one occasion and was accompanied by a higher rating for *anger*. The decision to reclassify *violent* in this way (in the absence of an appropriate entry in the thesaurus) therefore made no difference to the analysis.

Transformation of *other* responses (continued)

Response (*lexicon entry if different*)	Rational	Empirical	Relevant thesaurus entry(s)	Reclassification
disappointed	affective	affective	disappointed = sad	sadness
disappointment	affective	affective	disappointment = sadness	sadness
grief	affective	affective	grief = sadness	sadness
lonely	affective	affective	lonely = sad	sadness
unhappy	affective	affective	unhappy = sad	sadness
amazed	cognitive	affective	amazed = surprised	surprise
gobsmacked (*astonished*)*	cognitive	affective	astonished = surprised	surprise
pretend shock (*shock*)	affective	affective	shock = surprise	surprise
shock	affective	affective	shock = surprise	surprise
shocked	affective	affective	shocked = surprised	surprise
actor	-	-		neutral
aggressive	cognitive	external		neutral
agreeing	-	-		neutral
agreement	-	-		neutral
aloof	cognitive	cognitive		neutral
anticipation	-	-		neutral
apathy	affective	affective	apathetic = indifferent = neutral	neutral
attentive	-	-		neutral
bemused	-	-		neutral
blank	-	-		neutral
boredom	cognitive	physical/ bodily		neutral
cold	-	-		neutral

Gobsmacked*: "adj.** *Brit. informal* utterly astonished" (dictionary entry; OUP, 2002).

Transformation of *other* responses (continued)

Response (*lexicon entry if different*)	Rational	Empirical	Relevant thesaurus entry(s)	Reclassification
confused	cognitive	cognitive		neutral
confusion	cognitive	cognitive		neutral
consternation	-	-		neutral
contemplation	-	-		neutral
curious	cognitive	physical/ bodily		neutral
cynicism (*cynical*)	cognitive	cognitive		neutral
dazed	physical/ bodily	physical/ bodily		neutral
disagreement	-	-		neutral
disbelief	-	-		neutral
discomfort	-	-		neutral
disinterest	-	-		neutral
disparaging	-	-		neutral
doesn't understand	-	-		neutral
don't know look	-	-		neutral
doubt	cognitive	physical/ bodily		neutral
doubtful	cognitive	cognitive		neutral
dubious	-	-		neutral
expectant	cognitive	cognitive		neutral
expressionless	-	-		neutral
false	-	-		neutral
friendly	cognitive	external		neutral
gay	-	-		neutral
glazed	-	-		neutral
grimace	-	-		neutral
growl	-	-		neutral
illness	-	-		neutral

Transformation of *other* responses (continued)

Response (*lexicon entry if different*)	Rational	Empirical	Relevant thesaurus entry(s)	Reclassification
impassive	-	-		neutral
in thought	-	-		neutral
indifference	cognitive	Physical/ bodily		neutral
interest	cognitive	Physical/ bodily		neutral
interested	cognitive	cognitive		neutral
jolly	-	-		neutral
lack of interest	-	-		neutral
learning	-	-		neutral
lethargic	-	-		neutral
listening	-	-		neutral
lost	-	-		neutral
maybe	-	-		neutral
mistrust	-	-		neutral
misunderstanding	-	-		neutral
moody	-	-		neutral
musing	-	-		neutral
no man's land	-	-		neutral
nonplus	-	-		neutral
not amused	-	-		neutral
not sure	-	-		neutral
pain	Physical/ bodily	Physical/ bodily		neutral
perplexed	cognitive	cognitive		neutral
put on	-	-		neutral
puzzled	-	-		neutral
quizzical	-	-		neutral

Transformation of *other* responses (continued)

Response (*lexicon entry if different*)	Rational	Empirical	Relevant thesaurus entry(s)	Reclassification
received good news	-	-		neutral
reflecting	-	-		neutral
reflection on self	-	-		neutral
resignation (*resigned*)	cognitive	cognitive		neutral
silly	cognitive	external		neutral
skeptical	cognitive	Physical/ bodily		neutral
stoned	-	-		neutral
stressed	-	-		neutral
studious	-	-		neutral
subdued	-	-		neutral
superior	external	external		neutral
thinking	-	-		neutral
thoughtful	-	-		neutral
thoughtfulness	-	-		neutral
tired	Physical/ bodily	Physical/ bodily		neutral
ugly	-	-		neutral
uncertain	cognitive	cognitive		neutral
unknowing	-	-		neutral
unsure	-	-		neutral
unsure look	-	-		neutral
unsureness	-	-		neutral
watchful	-	-		neutral
weirdo	-	-		neutral
you what	-	-		neutral