

# Facial stereotypes and perceived mental illness

Charlie D. Frowd<sup>1</sup>, Sade Underwood<sup>1</sup>, Palwinder Athwal<sup>1</sup>, James M. Lampinen<sup>2</sup>, William B. Erickson<sup>2</sup>, Gregory Mahony<sup>3</sup> and John E. Marsh<sup>4</sup>

<sup>1</sup>Department of Psychology, University of Winchester, Winchester UK

<sup>2</sup>Department of Psychological Science, University of Arkansas, Fayetteville US

<sup>3</sup>Boston Police Department, Boston US

<sup>4</sup>School of Psychology, University of Central Lancashire, Preston UK

<sup>1</sup>Charlie.Frowd@winchester.ac.uk <sup>1</sup>S.Underwood.12@unimail.winchester.ac.uk <sup>1</sup>Palwinder.Athwal@winchester.ac.uk

<sup>2</sup>lampinen@uark.edu <sup>2</sup>wbericks@email.uark.edu <sup>3</sup>greg.mahoney@pd.boston.gov <sup>4</sup>JEMarsh@uclan.ac.uk

**Abstract**—It is well established that we carry stereotypes that impact on human perception and behaviour (e.g. G.W. Allport, “The nature of prejudice”. Reading, MA: Addison-Wesley, 1954). Here, we investigate the possibility that we hold a stereotype for a face indicating that its owner may have a mental illness. A three-stage face-perception experiment suggested the presence of such a stereotype. Participants first rated 200 synthetic male faces from the EvoFIT facial-composite system for perceived mental illness (PMI). These faces were used to create a computer-based rating scale that was used by a second sample of participants to make a set of faces appear mentally ill. There was evidence to suggest that the faces that participants identified using the PMI scale differed along this dimension (although not entirely as expected). In the final stage of the study, another set of synthetic faces were created by artificially increasing and decreasing levels along the scale. Participants were asked to rate these items for PMI and for six criminal types. It was found that participants assigned higher PMI ratings (cf. veridical) for items with inflated PMI (although there was no reliable difference in ratings between veridical faces and faces with decreased PMI). Implications of the findings are discussed.

**Keywords**—perceived mental illness; stereotype; victimisation; serious crime; EvoFIT

## I. INTRODUCTION

At the turn of the millennium, estimates were that a sizeable proportion of the adult population in Great Britain suffered from a mental illness, 16.3% in males and 19.4% in females (Singleton et al. [1]). More recent research reviews the prevalence of mental illness in specific occupations, especially jobs that are highly demanding, involve considerable emotional demands and have poor job security (Standfeld et al. [2]). In addition, the association between mental illness and criminal behaviour has received considerable research interest due to changes in public policy and increased media interest. From this research, it appears that vulnerable groups such as those with a mental illness tend to be over-represented in crime statistics. Furthermore, it seems that those with a mental illness are more likely to be victims of crime. For example, Hiday et al. [3] found that people in the United States with a severe mental disorder (e.g. schizophrenia and other psychotic illness) were more likely to be victims of crime than the general population. In a four-month survey, they found that 8.2% of their sample had been a victim of violent crime, which is over twice the national average. This is in contrast to the general tendency for members of the public to assume that

people with a mental disorder will be violent (Markowitz [4]). So, if those with a mental illness are more likely to be victimized, why is there a perception of this population as being more violent and why are they over-represented in the crime statistics (Fazel and Grann [5])?

It could be the case that people with mental illness simply do commit more violent crimes, or it could be they are more likely to be apprehended (Fazel and Danesh [6]). In particular, the police may be able to detect mentally-ill offenders more easily as perhaps they offend in public, behave in some way differently to non-mentally-ill offenders and/or do not flee the crime scene quickly (e.g. Hewitt [7]). However, could it be something about their physical appearance that makes them more likely to be apprehended?

The roots of such a suggestion date back to the time of Aristotle and the notion of physiognomy, the interpretation of a person’s character from his or her physical appearance, and then by several scholars in the nineteenth century. Charles Darwin suggested that physical cues can provide valuable signals for social interaction (for a discussion, see Valla et al. [8]). Cesare Lombroso noted the apparent (although not causal) relationship between biology and crime; for example, characteristics of murderers were found to include a strong jaw, large cheekbones and curly hair (Wolfgang [9]). Similar work was carried out by Francis Galton on the potential relationship between behaviour and shape of the human cranium and skeleton, and by Alphonse Bertillon who used a system of physical measurements and photographs to identify recidivist criminals (anthropometrics). Later, in the 1940’s, William Sheldon and colleagues proposed that extremes of body types, so called somatotypes, were associated with different types of crime. For example, mesomorphs, a muscular build with broad shoulders and narrow hips, were associated with that person committing violent crime. Their theory, along with researchers thereafter, while controversial, still appears to carry some weight, even though potentially more-effective measures of body assessment have now been proposed (see Maddan et al. [10]). Similarly, in 1969, Hooton [11] studied a large cohort of convicted persons in US prisons and revealed interesting correlations; for example, the height of a criminal was found to be positively related to tendency for murder but negatively related to propensity for committing property crime.

Relatedly, it is proposed that the human face is important for making judgements of personality and social behaviour (e.g. Allport [12]; Jost and Hamilton [13]; Watson [14]), in

particular for stereotyping. For example, there is considerable evidence that we have a stereotype for a physically-attractive face (for reviews, see Eagly, Ashmore, Makhijani and Longo [15]; and Fink and Penton-Voak [16]): possessing such a visage provides its owner with considerable benefit—a better-paid job, better salary, more-attractive partner, less severe punishment in court, etc. Indeed, research indicates that such individuals are less likely to be assigned a ‘criminal’ stereotype (e.g., Saladin, Saper and Breen [19]). In fact, this generalization has long been suggested (e.g. Bull and Green [17]), and some researchers (e.g. Clifford and Bull [18]) argue that it can exert a strong effect during criminal investigations—for example, by prompting members of the public to select someone from an identity parade because they have a ‘criminal appearance’. In general, it would appear that we consistently agree as to whether a face looks like a criminal (e.g. Bull and Green [17]), and this bias is strong enough that human observers are able to distinguish between faces of known criminals from non-criminal controls (Valla et al. [8]).

In the current study, we considered whether the human face, arguably our most identifying feature, could similarly convey information which hints at whether a person is suffering from a mental illness such as schizophrenia. If this were to be the case, then it is possible (although not inevitable) that we change our behaviour when interacting with someone with a mental illness, even leading to discrimination. In the current context, for example, offenders with such a condition could become a target for policing and therefore be subject to police apprehension more often.

In brief, our approach to answer this question was to ask people to make judgements of mental illness on a series of synthetic but realistic-looking male faces. Based on these data, we created two reference templates that were used as endpoints for a perceived-mental-illness (PMI) scale. Another group of participants were asked to use this scale to make further synthetic faces look ‘mentally ill’. Should there be evidence of a stereotype for mental illness, there should be a tendency for participants to favour the positive end of the scale. Next, another set of synthetic faces were manipulated towards either end of the scale and a third group of participants asked to rate them for PMI and a range of criminal types. Again, should the relevant stereotype exist, participants should accordingly tend to assign higher or lower PMI scores.

### A. Overview

So the main aim of the experiment was to investigate whether people have a stereotype for mental illness. In other words, is there evidence to support the notion that we hold a representation (in our cognitive system) for such a face? We reasoned that even though we may not have had direct experience of a person known [to us] to be mentally ill, various images and descriptions portrayed in the media may allow judgements of this type to be made; the same rationale applies to judgements of criminal types also collected (in Stage 3) of the experiment. A three stage approach was taken. In Stage 1, a group of 24 participants were given a collection of 200 computer-generated (synthetic) face images to rate for perceived mental illness (PMI). The 40 items that were rated lowest and highest were then used to create average templates, one for low and one for high PMI, respectively. These averages were then used as end points for a PMI scale; such scales have been created in a similar way in the past for a variety of global characteristics including honesty, health and attractiveness (see Frowd et al. [20]), and more recently for facial hardness, dominance, trustworthiness, and even for superficial characteristics such as level of suntan (Fig. 1). In Stage 2, a second group of participants were given a set of 18 randomly generated faces (with a known level of rated PMI from Stage 1) to manipulate along the scale, with the instruction to move the slider to produce a face that appeared to be mentally ill. In Stage 3, participants were given a further 18 faces to rate for PMI and six criminal types. All three stages revealed evidence of a facial stereotype for mental illness.

### B. Stage 1: Creation of low and high PMI reference averages Design

In Stage 1, participants were invited to rate a series of faces for perceived mental illness. Due to ethical concerns it was not possible to use photographs of real faces, and so face stimuli were generated from the EvoFIT facial-composite system. This forensic software produces computer-generated (synthetic) faces for witnesses and victims to create a face of a person they have seen to commit a crime; the resulting images are used by law enforcement to locate the relevant offender. In the current application, we made use of the numerous faces that are randomly generated by this system.

As is standard practice in psychological research, participants took part in an informed-consent procedure to decide whether or not to participate in each stage of the research. The informed consent involved, from the outset, an explanation of the general aims of the project, the time involvement and the (synthetic) nature of the face stimuli used.

#### *Participants*

Twenty-four undergraduates studying Psychology at the University of Winchester participated in the rating exercise. There were 22 females and two males and their age ranged from 18 to 49 ( $M = 21.0$ ,  $SD = 7.2$ ) years. Each person received course credit for their involvement in the study.



Fig. 1. Example reference templates (low-rated average on left, high-rated on right) for facial hardness (top row), trustworthiness (middle row) and suntan (bottom row).

### Materials

Two hundred faces from the EvoFIT composite system were randomly generated using the 40 year Western-European male database (Frowd et al. [20]). This model generates faces in the range 30 to 39 years which are similar to the photographic subjects involved in its construction. Each face was presented with a different hairstyle taken from an age-appropriate selection of approximately 250 alternatives. Images were printed in greyscale (the modality of the system)

to dimensions of 6 cm (wide) x 8 cm (high). Some images used in this stage of the experiment are presented in Fig. 2.



Fig. 2. Example randomly-generated faces used in Stage 1, for participants to rate for perceived mental illness.

### Procedure

Participants were tested individually (as was the case for all stages in the experiment). They were told that they would be presented with computer-generated faces for them to rate according to the degree to which each face appeared to be mentally ill (1 = not mentally ill ... 7 = mentally ill). It was emphasized that the faces to-be-seen were synthetic, and so not of actual people. Each face was presented in a different random order for each person and a rating was provided as requested. The exercise was completed in about 20 minutes, whereupon participants were debriefed.

### Results

The mean participant rating was calculated for each of the 200 randomly-generated faces; example images from this set are shown in Fig. 2 for a low-rated (left), a high-rated (right) and an intermediately-rated (middle) face. Reference templates for low and high PMI were then created using the procedure described by Frowd et al. [20]; this involved calculating image shape and texture averages for the 40 lowest- and 40 highest-rated items. By 'shape', we are referring to the physical location of facial features, their spacing on the face and the overall shape of the head; aspects of shape are particularly apparent between the two images shown in Fig. 3. For 'texture', we are referring to the greyscale intensity of pixels, which in groups represent the colour of the individual features (e.g. eyes and brows), skin tone, health of the face, etc. In Fig. 3, to assist with interpretation, we present a caricatured version of these averages, by exaggerating the shape information of each by 100% in both directions (see Frowd et al. [21] for a detailed description of how this procedure is achieved). As can be seen, there are obvious differences between reference templates, in particular since an increase in positive PMI rating appears to promote a face with greater facial weight and eyes that portray a less-intense gaze.



Fig. 3. Reference templates for the lowest- (left) and highest-rated (right) 40 randomly-generated faces. To illustrate differences, we have exaggerated (caricatured) the shape properties of each reference template from each other by 100%. The texture information in the face changes between images: notice, for example, the intensity of ‘bags’ under the eyes.

### Discussion

The production of reference templates by participant ratings indicates noticeable differences of PMI. In the next stage, we use these templates to create a psychological scale to allow synthetic faces to be changed or ‘manipulated’ along this dimension. The expectation was that participants would have a tendency to make faces appear more ‘mentally ill’—this is, towards a representation shown on the right in Fig. 3.

### C. Stage 2: Manipulation of faces by PMI scale

#### Design

In this stage, we developed a computer-based scale which would allow participants to alter the appearance of a synthetic face with the aim of creating a representation that appeared to be mentally ill. To do this, 18 faces were drawn from the original 200. Of these, six were selected to be the lowest rated PMI, six the highest rated, and six intermediately rated. Asking participants to manipulate these faces thus should result in moving them towards the positive PMI template for all items, but more so for faces rated as medium and low PMI.

In order to guard against participants simply moving the slider in one direction (i.e. to the right) to increase affect, the sense of the slider was randomized across items: half of the time positioning the slider to the right moved the face towards the positive reference average and the other half towards the negative reference average. The starting position of the slider was also randomized for each image (occurring anywhere along the scale in a random position). The slider scale had a total of 11 intervals.

#### Participants

Another 24 participants were drawn from the same participant pool and recruited in the same manner. There were 22 females and two males and their age ranged from 18 to 49 ( $M = 21.0$ ,  $SD = 7.2$ ) years.

### Materials

The PMI slider was implemented within the EvoFIT system. This decision was deliberate, partly since EvoFIT uses such sliding scales during face construction, for witnesses to improve the overall likenesses (Frowd et al. [20][22]). Example sliding scales (sliders) include facial health, honesty, masculinity and attractiveness (Frowd et al. [20]). Each scale changes the shape information of the overall face (shape and position of individual facial features) as well as the texture (greyscale pixel intensity) for the internal-features region (the central part including eyes, brows, nose and mouth); an example of a face manipulated along the PMI scale can be seen in Fig. 4. Another reason for manipulating the face in EvoFIT is that it could be used with witnesses and victims of crime as part of composite construction should it be found valuable in the current research. So, the reference templates were imported into EvoFIT, to allow the scale to be created, along with three sets of six faces (from the original 200) based on mean PMI ratings from the low, medium and high categories.

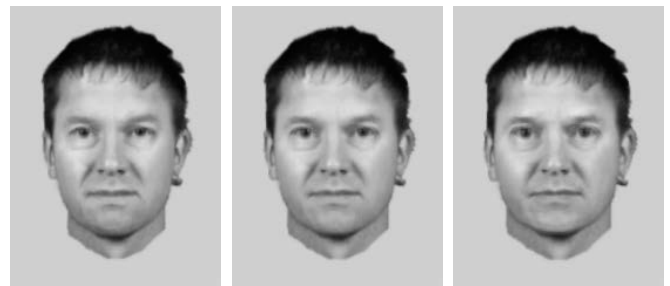


Fig. 4. An example randomly-generated face (centre) and resulting faces at the extreme negative (left) and positive (right) end of the PMI scale. In the study, there were 11 images in the sequence for participants to manipulate (i.e. there were an additional four steps between each pair of images shown here).

### Procedure

Participants were informed that a set of 18 randomly-generated faces would be seen and, for each, they should try to make the face appear mentally ill using the presented windows’ slider. Each face was presented sequentially and participants adjusted the slider as requested. The order of presentation of the 18 faces was random for each participant. The starting frame for each image (the position along the slider) and the slider’s sense were also randomised. The task took about 10 minutes to complete.

### Results

The scale position set by participants for each presented face was adjusted for sense of the scale (forward or reverse) and, for ease of interpretation, converted to percentage relative to the mid position—that is, from -100% (negative template) to 0 to +100% (positive template). Mean ratings by PMI category are presented in Table I. As can be seen, mean rating in the low and medium categories were very similar, but both of these values were much lower than mean rating of faces in the high category. Also, ratings in the high category were similar to zero, that is a setting corresponding to the centre position of the scale.

TABLE I. MEAN SCALE POSITION (IN PERCENTAGE) BY PMI CATEGORY

PMI Category <sup>a</sup>		
Low	Medium	High <sup>b</sup>
-22.4	-19.9	1.1
(9.2)	(9.2)	(9.2)

<sup>a</sup> Numbers in parentheses are standard errors of the individual response means produced from estimates in the GEE analysis (see accompanying text).

<sup>b</sup> Significantly greater than other two categories,  $p < .05$ .

Individual scale-position scores ( $N = 432$ ) from participants by PMI category were subject to Generalized Estimating Equations (GEE), a regression type analysis which essentially provides a combined by-participants and by-items analysis. The model used an ordinal scale with a logistic link function (as scores were Likert ratings), and participant number and PMI category were within-subjects' variables; an exchangeable structure for the Working Correlation Matrix and a model-based estimator were used. As was the case elsewhere in this paper, the data were initially checked for appropriacy for carrying out this kind of frequency-type analysis, and the resulting model was checked for sensible parameter values and standard errors (SE). PMI type emerged as a reliable predictor of scale position [ $X^2(2) = 13.6$ ,  $p = .001$ ] and parameter estimates (relative to low category) indicated that the low and medium categories were equivalent [ $B = 0.1$ ,  $SE(B) = 0.2$ ,  $p = .70$ ,  $Exp(B) = 1.1$ , 95% CI (0.8, 1.5)] but faces in the high category were assigned significantly more positive scale values than low [ $B = 0.6$ ,  $SE(B) = 0.2$ ,  $p = .001$ ,  $Exp(B) = 1.9$  (1.3, 2.7)]. Parameter estimates also revealed that values in the high category were reliably greater than those in medium category [ $B = 0.6$ ,  $SE(B) = 0.2$ ,  $p = .003$ ,  $Exp(B) = 1.8$  (1.2, 2.5)].

In a second analysis, we calculated whether scale position by category differed from the midpoint of the scale (i.e., from a rating value of 0). Mean participant scale position (averaged over six responses by PMI category) were analysed using a Wilcoxon Signed Ranks Test. Ratings in both low [ $Z = 2.0$ ,  $p = .05$ ] and medium [ $Z = 2.2$ ,  $p = .025$ ] categories were reliably different to zero, while ratings in the high category were not [ $Z = 0.2$ ,  $p = .81$ ]. The results suggest, therefore, that, overall, faces in the high PMI category were responded to as veridical (at the scale midpoint) while faces in the two other categories were adjusted towards the lower PMI scale reference.

### Discussion

Participants were invited to adjust the appearance of 18 faces (grouped into low, medium and high PMI category) to make them appear 'mentally ill'. The analyses revealed that faces in the high PMI category were not adjusted to be different from veridical, while faces in the low and medium category were assigned similar and lower ratings ( $M = \sim 20\%$ ). While the PMI scale does appear to be exerting a measurable effect on participant behaviour, its effect was not as intended; potential reasons for this finding are considered in the General Discussion. Next, we carried out a rating exercise on faces that were manipulated along the PMI scale.

## D. Stage 3: Face ratings

### Design

Another set of randomly generated faces along the PMI scale (in both positive and negative direction) were utilised. Participants rated them for PMI and a number of common criminal types (e.g. murderer, thief, burglar). Ratings of these criminal types somewhat mask the interest in perceived mental illness and this data could also be used to see if faces with artificially-inflated levels of PMI lead to an increase in rating not just for PMI, but also for ratings of criminal type. For the latter, it was anticipated that due to a tendency for people to associate mental illness with violent behaviour (e.g. Markowitz [4]), scale ratings would be higher for criminal occupations involving violent outcomes—murderer and rapist (cf. white collar criminal, thief, burglar and substance abuser).

In the experiment, participants rated six randomly-generated faces that had been manipulated by 100% in the negative direction along the PMI scale, a different six faces by 100% in the positive direction and a further six that had not been manipulated at all (veridical images). The design for image type (negative, neutral [veridical] and positive) was therefore within-subjects.

### Participants

Another group of participants were recruited from the same subject pool. There were six males and 24 females and their age ranged from 18 to 37 ( $M = 21.6$ ,  $SD = 4.9$ ) years. Participants were allocated to one of three testing sets with equal sampling.

### Materials

Eighteen faces were randomly generated using EvoFIT as in Stage 1. These items were manipulated by 100% in the negative direction along the scale and by 100% in the positive direction. Three testing booklets of 18 faces were created from these 54 images comprising six faces from each image type and rotating images around booklets. In this way, for whichever set was chosen, each participant was presented with one example of each identity and all three image types. Materials (Fig. 5) were printed on paper in the same way as in Stage 1.

### Procedure

Participants were asked to rate a series of faces along a number of dimensions. Faces were presented sequentially from one of three testing booklets and participants were asked to make a single rating (1 = low ... 7 = high) for each item. Once ratings had been collected, the same set of faces was presented for the second scale, and so forth for all scales. The rating scales used were PMI, murderer, rapist, thief, white-collar criminal, burglar and substance abuser. The rating order of scales was randomized across participants using a Latin-square design. Each participant received a different random order of presentation of images; each of the three testing sets was randomly selected with equal sampling. The rating exercise took approximately 30 minutes to complete.



Fig. 5. An example randomly-generated face (centre) used in Stage 3 that has been manipulated by 100% along the PMI scale towards the negative (left) and positive (right) reference template.

## Results

In the initial check of the appropriacy of the data for a frequency analysis, there were no actual observations for scale point 7 in the low category; this was overcome by changing one observation in this category for a rating of six to a rating of seven (to give a change with the least overall effect, in this case slightly increasing the central tendency in the low category); the resulting means by condition are presented in Table II. As can be seen, mean rating is similar in the low and medium categories, and both of these values are much less than scores in the high category.

TABLE II. MEAN RATING BY PMI CATEGORY

PMI Category <sup>a b</sup>		
Low	Medium	High <sup>c</sup>
2.7	2.6	3.2
(0.2)	(0.2)	(0.2)

<sup>a</sup> Numbers in parentheses are standard errors (see Table I notes).

<sup>b</sup> Rating scale extends from 1 to 7.

<sup>c</sup> Significantly greater than other two categories,  $p < .05$ .

The same GEE analysis from stage 2 was followed, except this time, the DV was participants' individual PMI ratings on a scale of 1 to 7 (rather than percentage scale position). GEE was significant for PMI [ $X^2(2) = 8.7, p = .013$ ], and parameter estimates (relative to the low category) indicated equivalence between low and medium [ $B = 0.1, SE(B) = 0.2, p = .79, Exp(B) = 1.1, 95\% CI (0.7, 1.5)$ ], but the high category was significantly greater than low [ $B = 0.5, SE(B) = 0.2, p = .015, Exp(B) = 1.6 (1.1, 2.4)$ ]. Parameter estimates also indicated benefit of high over medium categories [ $B = 0.5, SE(B) = 0.2, p = .007, Exp(B) = 1.7 (1.2, 2.6)$ ].

In a follow-up analysis, we considered the impact of PMI image category on the remaining rating scales—that is, for rapist, murderer, thief, etc. GEE were re-ran, except that type of rating scale (six scales) was added as a predictor along with image category (low, veridical and high). A significant interaction was predicted between these two factors, which was observed [ $X^2(20) = 104.0, p < .001$ ]. This interaction was then followed-up by separate analyses for each scale; these revealed that all scales were reliable [ $X^2(2) > 4.5, p < .1$ ] except for

burglar and white-collar criminal: high > (medium and low) ( $p < .1$ ) except that low and high were equivalent for rapist ( $p = .91$ ).

## Discussion

Participants were given 18 faces to rate along seven dimensions including PMI. The data indicate that participants rated faces in the high category reliably greater than medium or low; however, faces in the low and medium categories were treated equivalently. This same pattern of significant and non-significant effects was generally found to extend to ratings of rapist, murderer, thief and substance abuser.

## E. General Discussion

The aim of the study was to investigate whether there is evidence of a stereotype for mental illness. The approach taken was to ask participants to rate a set of randomly generated faces for this construct, to develop average (reference) templates from the lowest and highest rated sets of faces. There were clear differences in the appearance of the templates. In the next stage, participants manipulated randomly-generated faces along a PMI scale, with a view to making the face appear mentally ill, essentially moving between low and high reference templates. Unexpectedly, faces with medium and low ratings on PMI were manipulated towards the negative (cf. positive) template. In the final stage, randomly-generated faces were rated with higher PMI scores for faces which themselves had been manipulated along the PMI scale towards the positive sense.

It is clear from the findings that there does appear to be some kind of a facial stereotype for mental illness. The evidence for this suggestion is most evident in Stage 3: faces that had been shifted towards the positive (PMI) template were accordingly rated with higher scores for PMI (cf. randomly-generated faces that were unaltered). This result is key as ratings for a range of constructs were being collected, one of which was PMI—an approach to make the aim of Stage 3 less obvious. The results also indicated that faces manipulated towards the positive PMI template received reliably higher ratings for a range of criminal types (rapist, murderer, thief and substance abuser), indicating that people generally associate these criminal occupations with positive PMI. In other words, people with a perceived mental illness may also be associated with a stereotype of being a rapist, murderer, etc. This result is not unsurprising given the extensive literature which indicates the presence of strong stereotypes for a range of psychological constructs (e.g. Allport [12]).

Results from Stage 2 indicate that participants given high PMI categorical faces thought them to be no different from veridical (neutral). However, participants manipulated the faces in the two other categories to be *lower* (rather than *higher*) along this PMI scale. One possible explanation for this unexpected finding is that participants may have been reluctant to make this kind of judgement (and several of them commented as such) and so tended to use the scale in reverse—that is, in a more socially-acceptable way, to (in effect) make the face look more normal. Relatedly, it is also clear (Stage 3) that the scale is working appropriately when participants are asked to *rate* (rather than to *manipulate*) a given face.

When considering faces that had been manipulated [from veridical] towards the negative template, no reliable change [relative to veridical images] was found in Stage 2 (for setting along the slider) or in Stage 3 (for any of the seven rating scales). It may be that the lower half of the slider produces a representation which is too subtle to notice. Future research could explore whether moving the scale to a greater degree in the negative direction would have any measurable effect, basically then as a caricature (as illustrated in Fig. 3). Indeed, research could also investigate whether judgements of PMI might themselves be an exaggeration (a caricature) of the true effect (we argued above that the opposite effect, a so-called *anti*-caricature, may be occurring when *manipulating* a face). Future work could also consider whether a similar stereotype occurs for female rather than male faces, an issue of forensic importance given the increased interest in female offenders with mental illness (Fazel et al. [23]). Lastly, as suggested by one of the reviewers, it would be of interest to explore the extent to which our current results compare with perception of actual persons who are diagnosed with a mental illness. It is clear that we can distinguish criminal from non-criminal faces (Valla et al. [8]) and one would anticipate that, based on the current findings using photographs of synthetic faces, that we should be able to do the same using photographs of real faces.

In practical terms, greater positive affect was observed by presenting participants with faces manipulated along the PMI scale. In general, it would seem that the PMI scale would probably be useful for the EvoFIT system (e.g. Frowd et al. [24]), even though measured effect sizes ( $Exp(B)$ ) were fairly small but consistent (95% CI were reasonably narrow), along with other such global software tools (e.g. Frowd et al. [20]) for witnesses and victims to manipulate a face during forensic face construction. If a face had been created that eyewitnesses remarked could be improved by making it look more 'mentally ill', or even more like a rapist or murderer, then the scale could be of value forensically. This is currently being explored with witnesses and victims of crime in the UK and Europe.

In summary, the current study has provided evidence that we hold a stereotype for a 'mentally-ill' face. A scale that would appear to have value forensically was developed and this software tool is being currently used by police forces.

#### REFERENCES

[1] Singleton, N., Bumpstead, R., O'Brien, M., Lee, A., and Meltzer, H., "Psychiatric morbidity among adults living in private households", HMSO, UK, 2001.  
 [2] Stansfeld, S.A., Rasul, S.R., Head, J., and Singleton, N., "Occupation and mental health in a national UK survey", *Soc Psychiatry Psychiatr Epidemiol*, vol. 46, pp. 101–110, 2011.  
 [3] Hiday, V.A., Swanson, J.W., Swartz, M.S., Borum, R., and Wagner, H.R., "Victimization: A link between mental illness and violence", *International Journal of Law and Psychiatry*, vol. 24, pp. 559–572, 2001.  
 [4] Markowitz, F.E., "Mental illness, crime, and violence: Risk, context, and social control", *Aggression and Violent Behavior*, vol. 16, pp. 36–44, 2011.  
 [5] Fazel, S., and Grann, M., "The population impact of severe mental illness on violent crime", *American Journal of Psychiatry*, vol. 163, pp. 1397–1403, 2006.

[6] Fazel, S., and Danesh, J., "Serious mental disorder in 23,000 prisoners: a systematic review of 62 surveys", *Lancet*, vol. 359, pp. 545–550, 2002.  
 [7] Howitt, D., "Introduction to Forensic and Criminal Psychology" (4th edition). Harlow: Pearson Education Ltd., 2012.  
 [8] Valla, J.M., Ceci, S.J., and Williams, W.M., "The accuracy of inferences about criminality based on facial appearance", *Journal of Social, Evolutionary and Cultural Psychology*, vol. 5, pp. 66–91, 2011.  
 [9] Wolfgang, M.E., "Pioneers in Criminology: Cesare Lombroso (1825–1909)", *Journal of Criminal Law, Criminology and Police Science*, vol. 52, pp. 361–391, 1961.  
 [10] Maddan, S., Walker, J.T., and Miller, J.M., "Does size really matter: A reexamination of Sheldon's somatotypes and criminal behavior", *The Social Science Journal*, vol. 45, pp. 330–344, 2008.  
 [11] Hooten, E.A., "Crime and the man", New York: Greenwood Press, Publishers, 1968.  
 [12] Allport, G.W., "The nature of prejudice", Reading, MA: Addison-Wesley, 1954.  
 [13] Jost, J.T., and Hamilton, D.L., "Stereotypes in Our Culture", in *On the Nature of Prejudice: Fifty Years after Allport* (eds J.F. Dovidio, P. Glick and L.A. Rudman), Blackwell Publishing Ltd, Oxford, UK, 2005.  
 [14] Watson, D., "Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report", *Journal of Personality and Social Psychology*, vol. 57, pp. 120–128, 1989.  
 [15] Eagly, A.H., Ashmore, R.D., Makhijani, M.G., and Longo, L.C., "What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype", *Psychological Bulletin*, vol. 110, pp. 109–128, 1991.  
 [16] Fink, B., and Penton-Voak, I.S., "Evolutionary psychology of facial attractiveness", *Current Directions in Psychological Science*, vol. 11, pp. 154–158, 2002.  
 [17] Bull, R., and Green, J., "The relationship between physical appearance and criminality", *Medicine, Science, and the Law*, vol. 20, pp. 79–83, 1980.  
 [18] Clifford, B.R., and Bull, R., "The Psychology of Person Identification", Routledge: London, 1968.  
 [19] Saladin, M., Saper, Z., and Breen, L., "Perceived attractiveness and attributions of criminality: what is beautiful is not criminal", *Canadian Journal of Criminology*, vol. 30, pp. 251–259, 1988.  
 [20] Frowd, C.D., Bruce, V., McIntyre, A., Ross, D., Fields, S., Plenderleith, Y., and Hancock, P.J.B., "Implementing holistic dimensions for a facial composite system", *Journal of Multimedia*, vol. 1, pp. 42–51, 2006.  
 [21] Frowd, C.D., Bruce, V., Ross, D., McIntyre, A., and Hancock, P.J.B., "An application of caricature: how to improve the recognition of facial composites", *Visual Cognition*, vol. 15, pp. 1–31, 2007.  
 [22] Frowd, C.D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., McIntyre, A., and Hancock, P.J.B., "The psychology of face construction: giving evolution a helping hand", *Applied Cognitive Psychology*, vol. 25, pp. 195–203, 2010.  
 [23] Fazel S., Cartwright J., Norman-Nott A., and Hawton K., "Suicide in prisoners: a systematic review of risk factors", *Journal of Clinical Psychiatry*, vol. 69, pp. 1721–1731, 2008.  
 [24] Frowd, C.D., Skelton F.C., Hepton, G., Holden, L., Minahil, S., Pitchford, M., McIntyre, A., Brown, C., and Hancock, P.J.B., "Whole-face procedures for recovering facial images from memory", *Science and Justice*, vol. 53, pp. 89–97, 2013.