12-2011

Perceptions of Sexual Dangerousness: Accurate Identification of Sexual Offenders from Static Photographs

Amber Jean Culbertson-Faegre
University of Arkansas, Fayetteville

PERCEPTIONS OF SEXUAL DANGEROUSNESS: ACCURATE IDENTIFICATION OF SEXUAL OFFENDERS FROM STATIC PHOTOGRAPHS

PERCEPTIONS OF SEXUAL DANGEROUSNESS: ACCURATE IDENTIFICATION OF SEXUAL OFFENDERS FROM STATIC PHOTOGRAPHS

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts in Psychology

By

Amber J. Culbertson-Faegre
Missouri State University
Bachelor of Science in Psychology, 2008

December 2011
University of Arkansas

ABSTRACT

The present research expands understanding of the phenomenon of accurate identification of sexually dangerous men. Study 1 was designed to examine the influence of experience on accuracy of perceptions of sexual dangerousness. Receiving feedback about test trials increased accuracy on future trials at a marginally significant level. Study 2 was designed to determine the importance of specific facial features on these judgments. This study, however, failed to replicate the findings from the first study. Implications for Study 1, as well as possible explanations for Study 2 are discussed.

This thesis is approved for recommendation
to the Graduate Council.

Thesis Director:

_____

Dr. James M. Lampinen

Thesis Committee:

_____

Dr. Denise Beike

_____

Dr. Scott Eidelman

THESIS DUPLICATION RELEASE

I hereby authorize the University of Arkansas Libraries to duplicate this thesis when needed for research and/or scholarship.

Agreed _____
Amber J. Culbertson-Faegre

Refused _____
Amber J. Culbertson-Faegre

ACKNOWLEDGMENTS

Special thanks are due to Dr. James M. Lampinen who has allowed me the freedom to pursue this exploratory research. This thesis would not exist without the support, energy, and resources he has offered me the last three years. He also was kind enough to marry his wife Stephanie, who first suggested this research.

I am grateful for the many other people that helped support me through this process. Dr. Denise Beike and Dr. Scott Eidelman were kind enough to serve on my committee and offer me guidance and feedback. Kelly Degn, and many other research assistants, worked tirelessly to make sure the mechanics of these studies ran smoothly. Lindsey Sweeney is a true friend and has always been willing to lend a listening ear. Chris Peters has always offered inspiration for my research and Blake Erickson have always been quick to offer diversion when needed.

Over the last three years my entire family has been incredibly supportive of my academic goals. My younger brother, Justin Culbertson-Faegre, deserves special mention for acting as a sounding board, copy editor, crisis counselor, and reality checker. Many other friends have helped keep me sane, although Sam Aldridge's text message reminders to eat, sleep, and occasionally just go sit in the sun were especially useful.

Murphy, a huge old Great Pyrenees, offered me unconditional love until his death in January of 2010. Prophet and Oden continue to offer diversion and a reason to get up at 6:45 each morning in the way that only young hounds can. And Pterry, Quetzy, Mini-raptor and Darwin have acted exactly as you would expect chickens to act.

This thesis would not have been possible without funding from the Doctoral Academy Fellowship and the Howell's Fund.

TABLE OF CONTENTS

Perceptions of Sexual Dangerousness:

Accurate Social Cognition from Static Photographs

People make judgments about complete strangers on a daily basis, placing others into categories based only on the physical appearance of the target (Bruce & Young, 1998). In particular, humans rely heavily on the appearance of a target's face when judging personality characteristics (Macrae & Bodenhausen, 2000). Psychologists' understanding of what people are capable of accurately perceiving from only the face is growing. Researchers have previously demonstrated that people are capable of accurately judging the age, sex, and race of others based only on static photographs (e.g. Bruce & Young, 1998). Researchers also know that people are capable of accurately judging the sexual orientation of complete strangers, sometimes based on single facial features (Ambady, Hallahan, & Conner, 1999; Rule, Ambady, Adams, & Macrae, 2008; Rule, Macrae, & Ambady, 2009).

Sexual dangerousness, a propensity or history of committing sexually violent crimes, is a characteristic that poses a serious threat to children. People hold stereotypes and make judgments about a stranger's propensity to commit a variety of different crimes (Maclin & Maclin, 2004). Although people hold stereotypes about what criminals look like, and these stereotypes impact people's impressions of others, the accuracy of these stereotypes has received scant attention. Recent research by Carre, McCormic, and Mondloch (2010) showed the first empirical evidence that people can accurately judge a stranger's propensity for violence when only presented with a static photograph. No research to the best of my knowledge has focused on the accuracy of people's judgments of a stranger's sexual dangerousness.

Based on this void in the current body of research, I have begun to investigate the accuracy of perceptions of sexual dangerousness based only on static facial photographs. Preliminary work has found that people's subjective judgments of the "creepiness" of static photographs of a target correlates with the target's propensity to commit sexual assaults against children (Culbertson-Faegre, Lampinen, Sweeney, Erickson, & Peters, 2011). People are also able to accurately select men who have committed these sexual offenses, using only their static photographs.

The current studies were designed to investigate two important aspects of this phenomenon. The first goal was to evaluate the role that experience plays in these accuracy judgments. The second goal was to investigate the exact facial features that people rely on when making these judgments.

Impressions from Thin-Slicing

People form impressions of a stranger's personality even when exposed to very little information. Evidence suggests that these judgments are often quite accurate (Ambady, Bernieri, & Richeson, 2000). Thin slices of activity, or "excerpts of expressive behavior drawn from the ongoing behavioral stream," contain enough information to form highly accurate judgments about people (Ambady et al., 2000, p.203). The thin-slicing literature shows that people are able to determine internal states, personality characteristics, interaction motives, and social relations of strangers based on short videos of their behavior (see Ambady et al., 2000 for a full discussion). Even such ambiguous characteristics as the socioeconomic status and testosterone levels of strangers can be accurately judged by watching thin-sliced clips of behavior (Dabbs, Bernieri, Strong, Campo, & Milun, 2001; Kraus & Keltner, 2009).

Borkenau and Liebler (1993) found that personality traits could be accurately judged from thin-slices of behavior. They designed a study that investigated the relationship between self-evaluation of personality, evaluations by a person's intimate partner, and a stranger's evaluation of a target's personality. Borkenau and Leibler brought couples into the lab, and had them each complete a personality inventory for themselves and for their significant others. They then had each person read, on videotape, a standardized script. After watching these videotapes, strangers were asked to complete the same personality inventory for the target. Even when the tape was muted, Borkenau and Liebler found that strangers' evaluations of the people correlated with the self-evaluations and the evaluations of their partners. Borkenau and Liebler found that people could successfully predict the intelligence of a stranger when they were exposed to a 90 second standardized film clip, as long as audio was included.

In all of these studies, participants were given short video clips, sometimes accompanied by audio tracks. People are capable of accurately judging a great deal about strangers with a small amount of information. A growing body of evidence suggests that people can make accurate evaluations of others based on even less information. Face perception research focuses on accurate evaluations of others that can be made using only on static photographs of the face, or even specific facial features.

Facial Perception

People can determine a great deal of information from static photographs. The clearest information conveyed by the face is sex, which people can discern from standard monochromatic photographs (Brown & Perrett, 1993). People also form accurate impressions about a person's age and race based on their face (Bruce & Young, 1998). Specifically, participants can identify a

person's sex based on individual facial features, with the strongest cue being eyebrows, supporting earlier research that shows that the strongest cue in facial recognition is the eye and eyebrow region (Haig, 1986).

Accuracy for other impressions has recently come under investigation. Zebrowitz and Rhodes (2004) found evidence that intelligence and health can be predicted from static photographs of individuals in the lower half of the distribution of attractiveness. Brief impressions, based on static photographs, allow participants to predict if a male target is homosexual (Rule et al., 2008). Above chance predictions of sexual orientation were also made using only static images of faces, without the external cues like posturing or clothing. Rule and his colleagues tested the impact of the eyes and the mouths on accurate judgments of sexual orientation (Rule et al., 2008). They found that the judgments of sexual orientation were still above chance, indicating that these judgments are not based entirely on one single facial feature. They then tested the importance of salient facial features by presenting participants with photographs that had been cropped so that only the hair, eyes, or mouth were visible; all three presentations resulted in accuracy scores above chance. However, when photographs had all three of these salient features (i.e., hair, eyes, and mouth) removed, the accuracy of the judgments dropped to chance levels. This pattern of accuracy has been replicated when participants judge a female target's sexual orientation (Rule, Ambady, & Hallett, 2009).

This process appears fast, and category activation is at least partially automatic (Rule et al., 2009). The static photographs of homosexual and heterosexual males were presented as primes on a lexical decision task using words associated with the categories of "gay" and "straight." The primes facilitated response time for (respectively) straight and gay words,

suggesting that categorization of ambiguous group members is an automatic process. This research, however, does not address the possibility that people may be able to determine the criminal propensity or possible dangerousness of a stranger.

Dangerousness

Recently empirical research has begun evaluating people's judgments of faces as appearing dangerous. Most of this research measures the stereotypes that people hold about the facial structures of criminals. Research investigating this propensity to make judgments of people's criminality has shown that people form impressions of strangers based on photographs, although the validity of these impressions is often unclear (Maclin & Maclin, 2004).

Maclin and Maclin (2004) reported that people hold stereotypes about the facial features of specific types of criminals. Referred to as the "criminality" of a face, these features can decrease perceptions of attractiveness and increase the memorability of a face. However, this research did not evaluate the accuracy of perceptions of criminality, only people's perceptions of this characteristic.

Aggression, as a function of testosterone, can be predicted from the height-to-width ratio of male faces (Carre et al., 2009). Although the mechanism behind these judgments is structural it seems that people are unaware of the features they are relying on when making these judgments. Further research shows that people are capable of making judgments about the violence of an offender's conviction based only on static mug shots (Stillman, Maner, & Baumeister, 2010). Participants were presented with mug shots of sexual offenders and asked to estimate the aggressiveness of their crimes. Participants were able to make these judgments at levels above chance. Although this was not a direct measure of criminality, as all offenders had

committed sexual assaults, Stillman, et al.'s work is relevant because it shows that facial features or structures are used by people to make (accurate) judgments about a target's aggressive interpersonal behavior. Although this research speaks to the accuracy of people's judgments of violence, it does not comment on any judgments of an individuals' propensity or history of sexual assault. The current research was designed in an attempt to study the accuracy of individual's judgments in an attempt to evaluate perceptions of sexual offenders.

Over the last two years, I have conducted preliminary studies investigating the ability of people to judge propensity to commit sexual assaults based only on a static photograph (Culbertson-Faegre et al., 2011). Specifically, participants have been tested on their ability to judge whether a person has been convicted of a sexual assault against a child.

The initial study evaluated people's differential perceptions of men convicted of felony sexual offenses, compared to a control group of men convicted of a non-sexual felony crime. The entire study took place online. Undergraduate participants were presented with a series of mug shots from the Florida Department of Corrections. Half of the photographs (targets) were of white men that were convicted of sexual assaults of children under the age of 12, while half the photographs (distracters) were of white men that were convicted of a variety of non-sexual offenses. Each mug shot was presented alone on the screen. Participants were asked to rate the mug shots on a variety of characteristics, including measures of how "lonely", "deceptive," "happy," "weird," "angry," "trustworthy," and "creepy" the men appeared. These were traits that were initially used because of their potential relevance to sexual dangerousness.

Undergraduate participants judged sexual offenders to be significantly creepier, lonelier, and weirder than the non-sexual offenders. Participants also judged the sexual offenders to be

significantly less angry than the non-sexual offenders. This was, to my knowledge, the first empirical evidence that people are capable of making differential judgments of those with a history of sexual assault. The second study evaluated people's accuracy when selecting the photographs of men who had been convicted of sexual offenses against children. Undergraduate participants were presented with a forced judgment paradigm. On each trial, participants viewed mug shots of two men collected from the Georgia Department of Corrections. Each pair contained a photograph of a target white male who was convicted of a sexual assault of a child under the age of 12 and a distracter white male that was convicted of a variety of non-sexual offenses. The photograph pairs were matched for age. The photograph on the left was labeled as "A" and the photograph on the right was labeled as "B". Position of the target was counterbalanced across conditions. Participants were asked to select which man had been convicted of a sexual offense against a child, and to mark their answer on a six point confidence scale from 1 ("Very sure it is A") to 6 ("Very sure it is B"). Preliminary analysis suggested that participants were capable of accurately picking out which men had been convicted of a sexual offense based on the information presented in a standard mug shot.

The third study again evaluated people's accuracy when selecting photographs of men who had been convicted of sexual offenses against minors. This study addressed the situational cues of the previous studies by avoiding mug shots. Instead, the offender photographs were taken from PervertedJustice.org, a website that posts the photographs of men convicted of sexually soliciting a minor online. The photographs were the images that the targets sent to screen-names they believed belonged to minors. Only targets that had been convicted of sexually soliciting 12 and 13-year-old girls who were at least 10 years younger than the target

were included. The control photographs were aged matched white males that had included their photograph when posting solicitations for adult female sexual partners on the website Craigslist.com. Craigslist.com is an online classified website that includes dating and casual sex listings.

Undergraduate participants were presented with the same forced judgment paradigm as in the second study. The entire study took place online. Preliminary analysis suggested that participants were capable of accurately picking out which men had been convicted of a sexual offense using only facial photographs. This is the first evidence that people can accurately judge a person's propensity to sexually assault a child from static photographs of faces taken before a sexual offense occurred.

Statement of Problem

Although it has been established that people can sometimes distinguish between sexually dangerous men and distracters the exact mechanisms that allow this to occur are unclear. Two causal explanations for these mechanisms were tested in the current studies. First, the source of these judgments (innate or experience based) was investigated. Second, the individual features that people were using to make these judgments was investigated.

The first possible causal mechanism investigated was the source (innate or experience based) of these accurate judgments. If the accurate judgments of sexual dangerousness was innate, one would expect accuracy rates to be quite high. Otherwise, the adaptive advantage of the ability would be minimal. Previous research, however, indicates that the accuracy rates are small in absolute terms, although the differences have been consistently significant (Culbertson-Faegre et al., 2011). Therefore, it follows that these accurate judgments may be an experience-

based phenomenon, built up from news reports and other exposure to faces of sexual offenders. This would also fit with research that indicates that facial processing is an experience based phenomenon (Mondloch, Le Grand, & Maurer, 2002). This is true for both featural and configural facial processing techniques, although configural techniques take even more time to develop. It is possible that the previously demonstrated accuracy scores were a result of previous experience with sexually dangerous individuals. If this is the case it would be expected that additional experience should increase accuracy scores even more.

The experience given in this study was designed to optimally increase this phenomenon, if it is a result of experience. Cognitive skill learning, referring to implicit procedural learning that is also able to increase performance on future novel trials, requires experience (Gupta & Cohen, 2002). Other cognitive skill learning has been shown to be facilitated with feedback (Hattie & Timperley, 2007). Therefore, it is reasonable to predict task performance (accuracy) might increase with practice predicting sexual dangerousness as well.

Previous studies have evaluated the participant's experience as a victim of a sexual assault as well as the rate with which they watch television programs such as "To Catch A Predator" (Culbertson-Faegre et al., 2011). "To Catch A Predator" is a reality show based on sting operations where undercover civilians bust men who believed that they were going to meet with minors for sexual purposes. These factors have not been related the accuracy of an individual's judgments in the research reported earlier. It is unclear if the lack of relationship was because experience had no impact on accuracy or if our measures of experience were just poor. Also, there is a clear restriction of range when only inexperienced undergraduate participants are used. Experience range was expanded to evaluate the possibility that experience

with sexual predators is behind this phenomenon. In order to evaluate the impact of experience, accuracy scores were compared across groups with laboratory experience viewing faces of sexually dangerous men. Some participants took the test after no experience while others took the test after receiving training.

The second possible causal mechanism is that there are particular facial features that cue participants into the underlying characteristic of the person as a sexual predator. Research by Rule, et al. (2008) indicated that particular facial features convey more information about ambiguous group membership than others. In our previous research some participants mentioned that thin lips and beady eyes seemed to indicate that someone was a sexual offender. People exibit individual differences on physical characteristics such as facial features, facial feature relationships or spacing, and facial expressions. Which of these traits people are relying on when performing this task are unknown. Previous research on similar implicit social cognition has shown reliance on individual features, such as the demonstration that eyes alone convey implicit information about sexual orientation (Rule et al., 2008). Previous research has also demonstrated that some tasks are reliant on the relation between features, such as accurate perceptions of a stranger's propensity for aggression (Carre et al., 2009). Open-ended questions at the end of previous studies have shown that participants believe that they are relying on individual features,"beady eyes" and "thin lips", although that has yet to be empirically investigated (Culbertson-Faegre et al., 2011). The importance of individual features was tested in the current research by evaluating the accuracy of participants judgments when they only have specific facial features to rely on.

Study 1: The Impact of Experience and Feedback

The first study directly investigated the impact of experience on accurate judgments of a person's sexual dangerousness. Whatever cues these judgments are relying on may be learned. If these are learned judgments, one would expect that experience would improve accuracy, and a manipulation of experience should lead to an increase in the accuracy of participant's judgments. An online training program was created to present participants with a test at pre-test, a training module, and a test at post-test. The pre-test and post-test tests were identical in content. They each depicted 30 faces (15 targets and 15 distracters). The training module was designed to increase individual experiences with sexual offenders. In one condition participants experienced the rating task and received feedback on each judgment immediately. In another condition participants only experienced the rating task, without feedback. It was predicted that participants would make significantly more accurate judgments of a person's status as a sexual offender after completing the optimal training program. The independent variable of interest was the experience of judgment with feedback, the experience of judgment without feedback, or an alternate time-matched judgment. The dependent variable was the d-prime values on the test at pre-test and post-test.

Method

Participants. Undergraduate volunteers (n = 94, females = 58) from the University of Arkansas participated in exchange for partial completion of a course requirement. The average age of the participant was 19.21 years (SD = 1.46, Range = 18 to 27). The sample was predominately (87.23%) White with five participants that identified themselves as Asian, one participant that identified themselves as Black and one that identified themselves as Hispanic. Four participants indicated that they considered themselves to have multiple ethnicities.

Design. This study utilized a 3 (Condition: Judgment With Feedback, Judgment Without Feedback, or Alternate Judgment) x 2 (Test: Time 1 and Time 2) mixed-factor design with Condition being a between-subjects factor and Test being a within-subjects factor.

Materials.

Facial judgment test. The test consisted of a presentation of 30 faces, 15 target faces and 15 distracter faces. The static photographs of the sexual offenders (targets) and non-sexual offenders (distracters) were collected from the Florida Department of Corrections website. The targets were randomly selected White males who were convicted of the aggravated assault of a child under the age of 12. This was to be sure that the men had assaulted a child, since statutory rape laws vary so dramatically between states. The distracters were White males who were convicted of a non-sexual crime. In both cases, photographs of men wearing glasses were excluded. The distracters were age matched to the targets.

Participants were informed that, "Half of these men have been convicted of sexually assaulting a child." The participants were asked to make a judgment about whether the man in each photograph was a sexual offender. Each test photograph was presented individually. The photographs were presented in the center of the participant's screen, with an 8-point confidence scale from 1 ("Very sure it is a SEXUAL OFFENDER") to 8 ("Very sure it is a NON-SEXUAL OFFENDER") presented directly beneath the image.

This study moves to an individual presentation paradigm, from the previous forced choice paradigms, which addresses two concerns. First, the individual presentation paradigm bolstered the heuristic building function of the training by allowing for concise feedback to the participant about their choice. Second, the individual presentation paradigm also explicitly

ensured that people were able to make the judgments without having an alternative choice present.

Training materials. Training materials were presented between the pre-test test and the post-test test. The three exposure conditions (judgment with feedback, judgment without feedback, and a alternate judgment) received different training materials.

The judgment with feedback materials presented participants with a series of 150 mug shots, half of which were targets. The static photographs were presented individually, and the participant was instructed to make a judgment about each image by responding using the 8-point confidence scale described above. As soon as the participant made their judgment a second screen appeared, identifying the correct category. The participant then clicked to continue to the next trial.

The judgment without feedback materials were identical to the judgment with feedback materials, with one exception. The participant did not receive feedback on the correct category, but instead were sent directly to the next trial.

The alternate judgment condition was presented with photographs of 150 mug shots of men convicted of embezzlement. Participants were told that half of the men in the photographs were convicted of embezzlement. The static photographs were presented individually, and the participant was instructed to make a judgment about each image by responding on an 8-point confidence scale from 1 (“Very sure it is an EMBEZZLER”) to 8 (“Very sure it is NOT an EMBEZZLER”) presented directly beneath the photographs.

Demographics questionnaire. The demographics questionnaire looked at the participant's experience as a caregiver, authority figure, and victim (see Appendix A). The

questionnaire evaluated possible experience caring for children and interacting with sexual offenders. Questions about the participant's possible experiences with a victim or as a victim were included, and the participants rated the degree to which they could predict if someone is a sexual offender. The questionnaire also collected common demographic information.

Brief Mood Introspection Scale. The Brief Mood Introspection Scale (BMIS) is a measure of current mood included in Appendix B (Mayer & Gasche, 1988). This was included in order to determine if the conditions might have a different impact on mood, and to allow for exploration of the impact of mood on the effect.

Procedure. Study 1 was administered entirely online, using the Qualtrics website. Qualtrics is an online data collection website. Participants were randomly assigned to one of the counterbalanced versions of a presentation condition. Participants were informed that they would be completing a facial judgment task. They were informed that they would be taking two versions of a test and that their combined performance on the two tests could be rewarded with a 50 dollar gift card to the Razorback Store (a University store on campus). They were then presented with the facial judgment test at pre-test. After completing the test at pre-test, they were presented with one of the versions of the training materials. After completing the training materials they were reminded that the upcoming test was the one they may be rewarded with a gift card for completing. They were then presented with the test at post-test. They were then presented with the Demographics Questionnaire and a debriefing.

d' **Scoring.** The proportion of hits and false alarms were determined for both the pre-test and post-test. A hit was defined as correctly indicating that the person is a sexual offender when in fact the person is a sexual offender. A false alarm was defined as incorrectly indicating that a

person is a sexual offender when in fact the person is not a sexual offender. Based on the hit and false alarm data, the signal detection measures d' and β were computed (Snodgrass & Corwin, 1988). The signal detection measure provides a measure of sensitivity and thus provides an index of the degree to which participants can distinguish sexual offenders from non-sexual offenders. The signal detection measure β provides an index of response bias, with larger values indicating a bias against identifying that the person is a sexual offender. To provide a measure of the influence of training on sensitivity and response bias, pre-test/post-test difference scores were computed for both d' and the C Criterion, a measure of response strategy.

Results

Raw accuracy scores on the facial judgment task (total correct from 30 judgments) were evaluated for pre-test and post-test scores (See Table 1). Due to random sampling it was predicted that there would be no difference between conditions at pre-test. However, a one-way ANOVA showed that pre-test condition was related to participant accuracy on the facial judgment task, $F(2,93) = 4.447$, $MSE = 0.010$, $p = .014$. Specifically, the control condition ($M = 0.49$, $SD = 0.08$) performed significantly poorer than the feedback condition ($M = 0.57$, $SD = 0.10$) and the no feedback condition ($M = 0.54$, $SD = 0.11$), suggesting a failure of random assignment $F(1,91) = 8.037$, $MSE = 9.237$, $p = .006$. There was no significant difference in pre-test scores between the feedback condition and the no feedback condition ($t<1$).

It was predicted that condition (feedback, no feedback or control) would be related to the accuracy of the judgment tasks at post-test. A one-way ANOVA showed that condition was related to participant accuracy on the facial judgment task at post-test $F(2,93) = 6.104$, $MSE = 0.012$, $p = .003$. The feedback condition ($M = 0.61$, $SD = .11$) performed significantly better

than the no feedback condition (M = 0.54, SD = .11), $t(61) = 2.400$, $p = .019$. These results support the hypothesis that training, when accompanied by feedback, is enough to increase accuracy.

The average accuracy score, however, is bolstered by evaluating a potential shift in confidence and response criterion based on condition. The average d' and c criterion scores for pre-test and post-test were calculated. The average d' at pre-test, across conditions, was significantly higher (M = .171, SD = .559) than chance, $t(183) = 4.147$, $p < .001$. The average d' at post-test, across conditions, was also significantly higher (M = .2837, SD = .260) than chance, $t(183) = 6.517$, $p < .001$. There was also a significant difference between the post-test d' scores between conditions, $F(2, 91) = 4.927$, MSE = .331, $p = .009$. Participants in the judgment with feedback condition had higher d' scores on the post-test (M = 0.621, SD = 0.629) than participants in the judgment without feedback (M = 0..292, SD = 0.584) or the control conditions (M = 0.170, SD = 0.519) (see Table 1).

Accuracy on the intervening training materials was also investigated. There was no significant difference in accurate identification of targets between the experience with feedback (M = .370, SD = .122) and experience without feedback (M = . 444, SD = .224), $t(61) = -1.592$, $p > .05$. There was no significant difference in accurate identification of foils between the experience with feedback (M = .540., SD = .121) and experience without feedback conditions (M = . 503, SD = .206), $t(61) = .847$, $p = > .05$.

At this point a simple t-test of the difference scores between pre-test and post-test was planned. Due to a failure of randomization, however, the pre-test d' scores were significantly different between conditions, $F(2,91) = 4.623$, $p = .012$. In order to account for this variability, a

regression was used to control for the pre-test d' scores when analyzing the post-test d' scores. A one-way analysis of covariance (ANCOVA) showed that condition had a marginally significant effect on post-test d' scores, after covarying out pre-test d' scores, $R^2 = .250$, $F(1,91) = 3.362$, $p = .070$. Although the results were only marginally significant two orthogonal contrasts were created as exploratory analysis. The first contrast tested the prediction that feedback is a significant predictor of post-test d' scores. The second contrast tested the prediction that exposure to the sexual offenders alone is enough to significantly predict post-test d' scores. A simultaneous regression was conducted on the impact of the normalized pre-test d' scores, the orthogonal contrasts, and the interaction terms. The regression showed that these five factors predicted a significant proportion of variance in post-test d' scores, $R^2 = .128$, $F(5,88) = 2.588$, $p = .031$. Specifically, feedback predicted the post-test d' scores, $b = .297$, $t(88) = 2.943$, $p = .004$. Exposure to the images of the offenders alone, without feedback, did not predict post-test scores, $b = .081$, $t(88) = .813$, $p > .05$.

The plots of the target-distracter receiver operator characteristics (ROC) are shown in Figure 1. ROC curves visually communicate the degree to which one can distinguish between the two categories. The ability to discriminate between the two categories (targets and distracters) appears to be stronger for the targets when participants had experience with feedback. This suggests that, after making judgments about a stranger's sexual dangerousness and receiving feedback about the accuracy of those judgments, improved at recognizing future targets.

The responses to the demographics questionnaire showed a homogeneous sample with limited life experiences. Only one participant had worked as a law enforcement officer and

noone reported having spent time in prison or having experience as a jury member. Eight participants (8.5% of the sample) had interned or worked in the field of social work. Forty-eight participants had worked in a preschool or daycare setting (51.1% of the sample), although none of the participants had children of their own. Experience in preschool settings was not significantly correlated with pre-test accuracy scores, $r = .157$, $n = 93$, $p = .130$. The impact of condition on post-test d' scores with significant, even when the impact of participants having experience working with children was covaried out, $R^2 = .108$, $F(1,93) = 5.481$, $p = .006$.

Thirty-seven participants (41.5%% of the sample) reported watching fictional television shows that depict sexual offenders (e.g. Law and Order Special Victims Unit) on a weekly basis. Those participants reported watching an average of over six episodes a week ($M = 6.0428$, $SD = 7.651$). Self-reported viewing habits for fictional television shows that depict sexual offenders was not significantly correlated with pre-test accuracy scores, $r = .063$, $n = 93$, $p = .547$, Twenty participants (21.3% of the sample) reported watching non-fictional television shows that depict sexual offenders (i.e. "To Catch a Predator") on a weekly basis. Those participants reported watching an average of about five episodes a week ($M = 5.278$, $SD = 6.099$). Self-reported viewing habits for non-fictional televisions shows that depict sexual offenders was not significantly correlated with pre-test accuracy scores, $r = .124$, $n = 92$, $p = .240$.

Eighteen of the participants (9.8%) reported that they had been victims of a non-sexual crime in the past while 64 participants (34.8%) reported that they had at least one close friend or family member who had been the victim of a non-sexual crime. Five participants (2.7%) reported that they had been victims of a sexual crime while 55 participants (29%) reported that they had at least one close friend or family member that had been the victim of a sexual crime.

Of these four measures, the only variable that had any correlation with accuracy on the pre-test accuracy scores was personal history of having a friend or family member that was a victim of a non-sexual crime.

The vast majority of participants (169 participants, 91.8% of the sample) reported that they never knowingly interacted with sexual offenders during their work week. The other fifteen participants reported that they had only limited interaction with sexual offenders (M = 3.567, SD = 3.223, Range of 1 to 10). Participants were also asked to what extent they believed that people could predict sexual offenders, on a five-point Likert-type scale ranging from 1 ("Very Strongly Disagree") to 5 ("Very Strongly Agree"). Overall participants seemed neutral about their ability to predict who was a sexual offender (M = 2.92, SD = .92, range from 1 to 5).

Participants' scores on the BMIS were also computed based on the individual scales. The scores for each of these individual subscales are presented in Table 2. Overall participants rated their mood (M = 4.39, SD = 4.49) on the twenty point (-10 to +10) scale. Pearson's correlations were calculated to evaluate the relationship between post-test scores and each of the subscales of the BMIS. The Arousal/Calm subscale was significantly correlated with post-test performance, with people who were more aroused performing better on the post-test, r = -.262, n = 94, p = .011. The Positive/Tired subscale was correlated with post-test performance, but only to a marginally significant degree, r = -.192, n = 94, p = .064. No other subscales were correlated with post-test accuracy. A one-way analysis of covariance (ANCOVA) showed that condition still had a significant effect on post-test d' scores, after covarying out both Arousal/Calm and Positive/Tired subscales, b = .-.290, $t(93)$ = -2.883, p = .005.

Discussion

This study was designed to investigate the possible causal mechanisms of experience on these accurate judgments of the perceptually ambiguous category of sexual dangerousness. The hypothesis being tested was that experience has an impact on the accuracy of people's judgments of others as sexual offenders. The hypothesis supported, to some extent. Overall this study can be taken as a limited replication of the previous finding that participants are able to accurately identify sexual offenders at levels above chance. This study also demonstrates that participants in the judgment with feedback condition scored marginally higher than the participants in the judgment without feedback and alternate judgment conditions.

Participants who made judgments about sex offender status and received feedback, controlling for initial ability to discriminate between the faces, outperformed participants in the other two conditions. This suggests that experience can increase accuracy of these judgments of sexual dangerousness. Demonstrating that experience influences accuracy when predicting sexual dangerousness is important because a fluctuation suggests that it is not a purely innate or static phenomenon. It is something that can be learned, and an effect size that can be moved around. It also suggests that there may be natural individual differences (based on overarching life experiences) between people.

Study 2: Facial Regions

Study 2 was designed to investigate which facial features people rely on when judging if a person is potentially sexually dangerous. Many of the cues people rely on to determine age, sex, and race/ethnicity come from the eyes of a target (Bruce & Young, 1998). Ambady and Rule (2008) have shown that people are able to accurately judge the sexual orientations of both males and females based only on static images depicting the eyes of a stranger (Rule et al., 2008; Rule

et al., 2009). The eyes, in particular, seem to convey information concerning personal characteristics.

Because eyes seem to communicate a great deal of information to the observer, it was hypothesized that this may also be the facial region that communicates a person's status as a sexual predator. It was expected that a photograph of the entire head or the entire face provide more information than the eyes alone, but that the eyes alone would be sufficient to produce differential selection of the sexual offenders. To test this hypothesis mug shots were cropped to present individual facial features. Photographs of targets and distracters were presented in four different forms: the entire head, the face with hair and frame removed, just the eyes, and just the mouth. The accuracy of judgments of the sexual offenders was compared between the cropping conditions.

Method

Participants. Undergraduate volunteers (n = 79, females = 52) from the University of Arkansas participated for partial completion of a course requirement. The average age of the participant was 19.32 years (SD = 1.46, Range = 18 to 26). The sample was predominately (95%) white with two participants that identified themselves as Black and one that identified his or herself as Hispanic. One participant indicated that he or shenoone had multiple ethnicities.

Design. Presentation type (Presentation: Head, Face, Eyes, Mouth) was a within-subjects manipulation.

Materials.

Facial Judgment Task. The forced judgment task was the same forced judgment paradigm as previously used in the research lab (Culbertson-Faegre et al., 2011). Static

photographs of the sexual offenders (targets) and non-sexual offenders (distracters) were collected from the Florida Department of Corrections website. The targets were randomly selected White males who were convicted of the aggravated sexual assault of a child under the age of 12. Photographs of men wearing glasses were excluded in order to allow a clear view of the facial features. The distracters were White males who were convicted of a non-sexual crime The distracter photographs were age matched to the target photographs.

The photographs were cropped to create four different presentations. The head crop included the entire mug shot, including the background and hair details. The face crop showed an elliptical crop of the target's face, from the top of the forehead to the bottom of the chin. Ears and hair were not included in the image. The eyes crop showed a rectangular crop of the target's eyes, excluding the eyebrows of the target. The mouth crop showed the half ellipse including the mouth and chin of the target (see Figure 2). These cropping methods were loosely based on the guidelines set by Rule and his colleagues (Rule et al., 2008).

Forty photograph pairs, each including a target and a distracter, were presented. The 40 photograph pairs were randomly divided into four sets of ten. One set of photographs was presented in the head crop format. Another set of photographs was presented in the face crop format. The third set of photographs was presented in the eye crop format. The fourth set of photographs was presented in the mouth crop format. Across participants, each of the four sets of photograph pairs were presented equally often in each of the four cropping formats. The pairs of age-matched static photographs of White men were presented side by side. The photograph on the left was labeled "A", while the photograph on the right was labeled "B". In addition to counterbalancing the cropping format across pairs of pictures, within each of the four sets of 10

photograph pairs, the target photograph appeared in the 'A' position five times and the 'B' position five times. Across participants, materials were counterbalanced such that each picture occurred equally often in the "A" and "B" position.

Participants were asked to select which man had been convicted of a sexual offense against a child, and to mark their answer on an eight point confidence scale from 1 ("Very sure it is A") to 8 ("Very sure it is B"). The trials were presented in a single randomized order. The location of the sexual offender, in position "A" or "B" was counterbalanced across participants. In addition, for each participant the sexual offender was positioned equally often in the A and B position for each of the four types of presentations.

Demographics Questionnaire. The demographics questionnaire was identical to the one used in Study 1.

Brief Mood Introspection Scale. The BMIS was identical to the one used in Study 2.

Procedure. Participants were recruited from the undergraduate psychology experiment pool at the University of Arkansas. The experiment was administered entirely online, using the Survey Monkey website. Survey Monkey is a commercial data collection website. Participants were randomly assigned to one of the counterbalanced versions of a presentation condition. They were directed to a website where they were presented with a consent form informing them of their rights as participants. Participants were informed that they were completing a face evaluation task. They were then presented with the face evaluation task and the demographics questionnaire. When they are finished, they were then presented with a debriefing.

Scoring. For each participant, an accuracy score was determined for each of the four presentation conditions. Accuracy in each of the four conditions was determined in two ways.

First, for each of the four within participant conditions, the proportion of correct responses was determined. In addition, a weighted accuracy score was also calculated within each of the four conditions, which weighed more heavily the responses given with higher levels of confidence.

$$Accuracy_{weighted} = \frac{\sum_{i=1}^{10} c_i a_i}{\sum_{i=1}^{10} c_i}$$

To calculate the weighted accuracy score, each response was assigned a value of a_i=1 (accurate) or a_i=-1(inaccurate) and this value was multiplied by the participant's confidence (c_i) between 1 and 4 indicating the level of confidence (e.g., 4=very confident, 3=moderately confident, etc.). Within each condition, these values were summed and divided by the sum of the confidence scores.

Results

Both the proportion correct and the weighted accuracy scores for the different presentation conditions were compared against chance using single-sample t-tests. First the proportional accuracy scores were compared to chance for each of the four conditions. When participants saw the entire head they were accurate at levels (M = .487, SD = 0.144) no different from chance, t(78) = -.784, p = .436. When participants saw only the cropped face they were accurate at levels (M = .520, SD = .162) no different from chance, t(78) = 1.121, p = .266. When participants saw only the cropped mouths they were accurate at levels (M= .471, SD = . 166) no different from chance, t(78) = -1.573, p = .120. When participants saw the cropped eyes they were accurate at levels (M = 0.537, SD = .020) marginally different from chance, t(78) = 1.882, p = .064. This means that the raw accuracy data suggests that participants performed at levels no different from chance when presented with the heads, faces, and mouths of participants.

Even the accuracy demonstrated in the eyes condition was only marginally different from chance.

The weighted accuracy scores were also compared to chance performance levels (See Table 3). When participants saw the entire head they performed at levels (M = -0.163, SD = 0.377) significantly worse than chance, $t(78) = -3.826$, $p < .001$. When participants saw the cropped faces they performed at levels (M = -0.0952, SD = 0.431) marginally below chance, t (78) = -1.966, $p = .053$. When participants saw the cropped mouths they performed at levels (M = -0.236, SD = 0.501) significantly below chance, $t(78) = -4.175$, $p < .001$. When participants saw the cropped eyes they performed at levels (M = -0.419, SD = .047) significantly lower than chance, $t(78) = -2.367$, $p = .020$.

The responses to the Demographics questionnaire showed a homogeneous sample with similar life experiences to the sample used in Study 1. Two participants had served on a jury. Nobody had experience working as a law enforcement officer or as a prison inmate. Only 2 (2.5% of the sample) participants had interned or worked in the field of social work, although 17 (21.5%) had some experience caring for children in a childcare or school setting.

Fifteen participants (19% of the sample) reported watching television shows that depict sexual offenders (i.e. “To Catch a Predator”) on a weekly basis. Those participants reported watching an average of about two episodes a week (M = 1.867, SD = 1.50). Forty-four participants (55.7% of the sample) reported watching fictional television shows that depict sexual offenders (i.e. “Law and Order SVU”) on a weekly basis. Those participants reported watching an average of two episodes a week (M = 2.114, SD = 1.58).

Five of the participants (6.3% of the sample) reported that they had been victims of a non-sexual crime in the past while 25 participants (31.6%) reported that they had at least one close friend or family member who had been the victim of a non-sexual crime. Five participants (6.3%) reported that they had been victims of a sexual crime while 28 participants (35.9%) reported that they had at least one close friend or family member that had been the victim of a sexual crime.

The vast majority of participants (97.5% of the sample) reported that they never knowingly interacted with sexual offenders during the course of their job. The other two participants reported that they had varied interaction with sexual offenders (M = 11, SD = 11.314, Range of 4 to 20). Participants were also asked to what extent they believed that people could predict sexual offenders. They were given a five-point Likert-type scale ranging from 1 ("Very Strongly Disagree") to 5 ("Very Strongly Agree"). Overall participants seemed neutral about their ability to predict who was a sexual offender (M = 2.557, SD = 1.095, range from 1 to 5).

Participants' scores on the BMIS were also computed based on the individual scales. Overall participants rated their mood (M = 5.981, SD = 2.307) on the twenty-point (-10 to +10 scale) provided.

Discussion

The results reported above are in direct contradiction to the hypothesis that accuracy scores would vary based on the cropping of the photos. These results also contradict the main effect results for several previous studies, including the pattern of results demonstrated in Study 1 of this paper. There are two possible explanations for these data. The first is that it is an

accurate depiction of people's abilities to accurately identify sexual offenders. The second is that an aspect of this specific paradigm disrupted the previously demonstrated phenomenon.

If these data are assumed to be accurate there is a chance that it disproves the previous research (Culbertson-Faegre et al., 2011). Because the finding of previous studies have been so consistent it is important that the current (oppositional) results are verified before discarding this line of research. There is also a chance that the particular question that was asked here, or the way that it was investigated, disrupted the previously demonstrated accuracy rates.

The most likely reason for these results is that the task itself was so difficult or distracting that people performed, overall, at chance or even below chance. The performances on the head and face crops are most worrisome because they should replicate previous findings. The only thing about this paradigm that is different from previous studies is that the cropped images of the eyes and faces are distributed throughout the trials. That may have disrupted holistic processing of the faces, or decreased attention given to the task in some way. Previous research on holistic processing has shown that asking people to focus on individual features, or having them verbally describe individual features, is enough to disrupt future holistic processing (Navon, 1977; Schooler & Engstler-Schooler, 1990). It is possible that the feature trials acted as a similar prompt to focus on featural processing.

General Discussion

People make fast accurate judgments about strangers that they meet, based only on their faces (Rule et al., 2009). These judgments include intuitively "obvious" characteristics: gender, age, and ethnicity (Bruce & Young, 1998). They also include ambiguous characteristics: political affiliation, personality traits, and sexual orientation (Naumann et al., 2009; Rule et al., 2009).

Researchers are also beginning to suggest that people are capable of determining how dangerous a stranger is (Carre et al., 2009). Research in this lab has supported the hypothesis that people are capable of accurately identifying men that have a history of dangerous sexual assaults (Culbertson-Faegre et al., 2011).

The current research was designed to explore two different mechanisms that could impact these accurate judgments of sexual dangerousness. Study 1 was designed to test the importance of experience on accuracy. Study 2 was designed to test the degree to which participants relied on individual facial features when making these judgments.

Study 1 was designed to investigate the possibility that experience increases the accuracy of sexual dangerousness. In this study experience was manipulated experimentally. The largest increase in accuracy occurs when participants are asked to make a great number of judgments and feedback is given immediately after each judgment. This suggets that accuracy is at least partially a function of experiences. These results are especially remarkable considering that the participants in the judgment with feedback condition only had experience with feedback for a short period of time between the two testing periods.

Testing the influence of experience on accuracy is a relatively novel approach in itself. Basic configural feature processing seems to require time and experience to build (Mondloch, Le Grand, & Maurer, 2002). However, the facial perception literature has not focused on the impact of experience on accurate judgments of strangers. The cognitive literature on facial processing, however, has shown that the way that people process and remember faces may be a function of their experience processing faces. Specifically, it seems that holistic processing may increase with experience with a type of object (Gauthier, Curran, Curby, & Collins, 2003). Holistic

processing of faces, and even superior holistic processing for specific sorts of faces, is likely a result of increased experience. If this is true, it suggests that , if the accurate identification of sexual dangerousness increases with experience, it might be a result of increased holistic processing. Future research should be designed to investigate the importance of holistic processing on this task by disrupting holistic processing and determining if the phenomenon disappears.

Study 1 demonstrated that experience, accompanied with feedback, was enough to increase accuracy, at a marginally significant level. This increase in accuracy, in some ways, poses more questions than it answers. If are able to increase accuracy with this limited experience, it begs the question of the potential limits of experience. Research could determine if there are potential ceiling effects of experience, and how hard is it to train someone to make judgments at optimal performance. The previous studies have always used inexperienced undergraduate participants. This study will be replicated with more participants, since the results of this study are only marginally significant. Research is currently being conducted to determine the maximum impact that experience can have on these judgments. Specifically, I am interested on the impact that a career spent interacting with sexual offenders (such as forensic psychologists) has on the accuracy of these judgments.

Study 2 was designed to investigate the possibility that there are particular facial features that participants rely on when judging sexual dangerousness. Participants made judgments relying on the four different types of cropped photograph presentations. Due to a failure to obtain a significant raw proportion of correct responses, these results fail to address the question the importance of facial features in the previous phenomenon. Study 2 failed to replicate the

accuracy scores previously demonstrated, even the results described in Study 1. Future research is being planned that will address this question. Three different methodologies are under consideration, each addressing a different possible explanation for the results of Study 2.

The first possibility is that the paradigm used was too complex, and disrupted the facial processing of participants entirely. The task was within-participants, with a mixed trial presentation. The featural cropping may have focused their attention on individual features. The cropped photos from Study 2 will be used to run a second study, this time using a between-participants design. If a main effect, at least for the head and face conditions, is demonstrated it would indicate the results of Study 2 were a fluke of the paradigm used. If no main effects, even in the head and face conditions are demonstrated again, it would suggest that the entire phenomenon will need to be reinvestigated. The second experiment that is being considered would, instead of controlling what the participants see, measure what the participant focuses on. By using an eye-tracker or mouse-tracker participants could focus on the portions of the face that they believe will help them. Raw and proportional data would demonstrate what participants are choosing to look at while a correlation between featural focus and accuracy scores could speak to the actual impact of the features.

The third experiment that is being considered would remove the variability of participants entirely. The quantitative characteristics of previously tested photos would be evaluated. Any difference between these two types of photos could speak to the facial features that participants are relying on. Several methods of obtaining this information are being considered. Raw measurements of the facial features and relationships between the features have been used in previous studies (Bruce & Young, 1998). A new approach, known as principal component

representation analysis (PCA), would also allow this task to be approached statistically. PCA takes a large number of faces and attempts to define an average face, while keeping as much information about high variance regions as possible (Hancock, Bruce, Burton, 1997). This technique has previously been used to produce prototypical faces, but it might allow for a straightforward and intricate evaluation of the photographs without having to involve participants. Similarly, a holistic creepiness dimension could be created and evaluated using facial construction software such as evoFIT (Frowd et al., 2008).

Overall, the implications of this research on the general understanding of social cognition and facial perception are remarkable. The degree to which this effect can be moved or manipulated communicates the overall importance of this finding. Already it has demonstrated that people are capable of determining much more about a stranger's propensity (or history) of sexual dangerousness than previously expected. If future research is able to increase accuracy or determine the specific features that these judgments are relying on, it could help direct the wider research on sexual dangerousness.

The thin-slicing and facial processing literature has continued to grow in recent years. This newly understood method of social cognition has continued to grow and theoretical explanations for this phenomenon are vague and incohesive. It is important that researchers continue to stretch the limits of what personality traits and behaviors can be perceived with limited facial information, as well as understanding the paths through which these social perceptions are formed and developed is important. This research (at least Study 1) further bolsters the expansion of the facial perception research to include sexual dangerousness. It also

expands the theoretical understanding of this phenomenon by demonstrating that the accuracy of these judgments are based on experience.

Applications for this research present themselves in both applied and theoretical settings. There may be applications for this research in helping protect potential victims or professionals who are attempting to protect children from sexual assault. If particular facial features were picked out, high-risk victims could be taught to watch for those features, or they could complete training to increase the likelihood that they could avoid offenders.

There is currently debate among forensic psychologists about what causes sexual dangerousness. There is an understanding that a propensity to commit sexual assault arises as result of life experiences, although genetic factors that influence brain development may also be involved (Ward & Beech, 2005). Determining what differentiated sexually dangerous targets it might help further an understanding of the factors that lead to sexual dangerousness. If the facial cues that were important were malleable (reliant on expression, musculature cues, etc.) it would suggest that sexually dangerous behavior might be a result of life experiences. If the facial cues that were important were static (feature type, facial width, holistic processing) it would suggest that predisposition to act in sexually dangerous ways might be set early in life. Knowing what factors increase risk of sexual dangerousness may help the field to fomulate more productive treatments or interventions for known or high-risk sexual offenders.

References

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology, 32*, 201-271.

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology, 77*, 538-547.

Borkenau, P. & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology, 65*, 546-553.

Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception, 22*, 829-840.

Bruce, V., & Young, A. (1998). In the eye of the beholder: The science of face perception. Oxford University Press, New York.

Carre, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20*, 1194-1198.

Culbertson-Faegre, A., Lampinen, J. M., Sweeney, L., Erickson, B., Peters, C. S., (2011, March). *Bad apples versus creepy apples: Differential impressions of static photos of sexual offenders*. Poster session presented at the American Psychology-Law Society International Congress on Psychology and Law, Miami, FL.

Dabbs, J. M., Bernieri, F. J., Strong, R. K., Campo, R., & Milun, R. (2001). Going on stage: Testosterone in greeting and meetings. *Journal of Research in Personality, 35*, 27-40.

Frowd, C.D., Bruce, V., Change, H.Y., Plenderleith, Y., McIntyre, A.H. & Hancock, P.J.B. (2008). Predict your child: A system to suggest the facial appearance of children. *Journal of Multimedia, 3*, 28-35.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003) Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6*, 428-432.

Gupta, P., & Cohen, N. J. (2002) Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological Review, 109,* 401-448.

Haig, N. D. (1986). Exploring recognition with interchanged facial features. *Perception, 15*, 235-247.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (1997). Testing principal component representations for faces. In, J.A. Bullinaria, D.W. Glasspool & G. Houghton (Eds) *Proceedings of 4th Neural Computation and Psychology Workshop.* (pp. 84-97). London: Springer-Verlag.

Hattie, J., & Timperley, H. (2007) The power of feedback. *Review of Educational Research. 77,* 81-112.

Kraus, M. W., & Keltner, D. (2009). Signs of socioeconomic status: A thin-slicing approach. *Psychological Science, 20*, 99-106.

Lampinen, J. M., & Culbertson-Faegre, A. (2009). *Accurate judgment of sexual offense history from static photographs*. Manuscript in preparation.

MacLin, O. H., & MacLin, M. K. (2004). The effect of criminality on face attractiveness, typicality, memorability and recognition. *The North American Journal of Psychology, 6*, 145-154

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*, 93-120.

Mayer, J. D., & Gaschke, Y. N. (1988). The experience and meta-experience of mood. *Journal of Personality and Social Psychology*, 55(1), 102-111.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31,* 553-566.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin, 35.* 1661-1671.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. Cognitive Psychology, 1977, 9, 353-383.

Rule, N. O., Ambady, N., Adams, R. B., & Macrae, C. N. (2008). Accuracy and Awareness in the Perception and Categorization of Male Sexual Orientation. *Journal of Personality and Social Psychology, 95*, 1019-1028.

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology*.

Rule, N. O., Macrae, C. N., & Ambady, N. (2009). Ambiguous group membership is extracted automatically from faces. *Psychological Science, 20,* 441-443.

Schooler, J.W., Engstler-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology, 22*, 36–71

Snodgrass, J.G. & Cowin, J. (1988). Pragmatics of measuring recognition accuracy: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34-50.

Stillman, T. F., Maner, J. K., & Baumeister, R. F. (2010). A thin slice of violence: Distinguishing violent from nonviolent sex offenders at a glance. Evolution and Human Behavior, 31, 298-303.

Ward, T., & Beech, A. R. (2005). An integrated theory of sexual offending. *Aggression and Violent Behavior, 11*, 44–63.

Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to “bad genes” and the anomalous face overgeneralization effect: Cue validity, cue utilization, and accuracy in judging intelligence and health. *Journal of Nonverbal Behavior, 28*, 167-185.

Table 1. *Raw Accuracy and D-prime Scores from Study 1*

Condition	Pre-test Raw		Pre-test D’		Post-Test Raw		Post-Test D’	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
FB	0.570	0.100	0.4096	0.572	0.615	0.110	0.622	0.630
NFB	0.543	0.114	0.235	0.585	0.548	0.110	0.291	0.575
Cntrl	0.494	0.087	-0.011	0.446	0.519	0.105	0.170	0.519
Total	0.535	0.105	0.208	0.560	0.559	0.114	0.353	0.599

Table 2. *BMIS Scores by Condition from Study 1*

Condition	Pleasant		Arousal		Positive		Negative		Overall	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
FB	1.793	4.118	15.276	3.421	5.586	2.639	6.724	2.631	14.241	4.120
NFB	3.910	5.680	16.029	3.572	6.912	3.223	6.177	2.691	14.853	4.730
Cntrl	3.936	6.055	17.194	3.209	7.355	3.094	6.903	2.773	14.032	4.665
Total	3.266	5.418	16.181	3.461	6.649	3.068	6.5851	2.690	14394	4.494

Table 3. *Raw Accuracy Scores for Study 2*

Condition	*M*	*SD*	*t*	*df*	*p*
Eyes	0.537	0.173	1.882	78	.064
Face	0.520	0.161	1.121	78	.266
Head	0.487	0.143	-.784	78	.436
Mouth	0.471	0.166	-1.572	78	.120
HeadFace	0.503	0.106	.246	78	.807
EyesWeighted	-.112	.419	-2.367	78	.020
FaceWeighted	-.088	.428	-1.831	78	.071
HeadWeighted	-.163	.378	-3.827	78	.001
MouthWeighted	-.084	.245	-3.045	78	.003

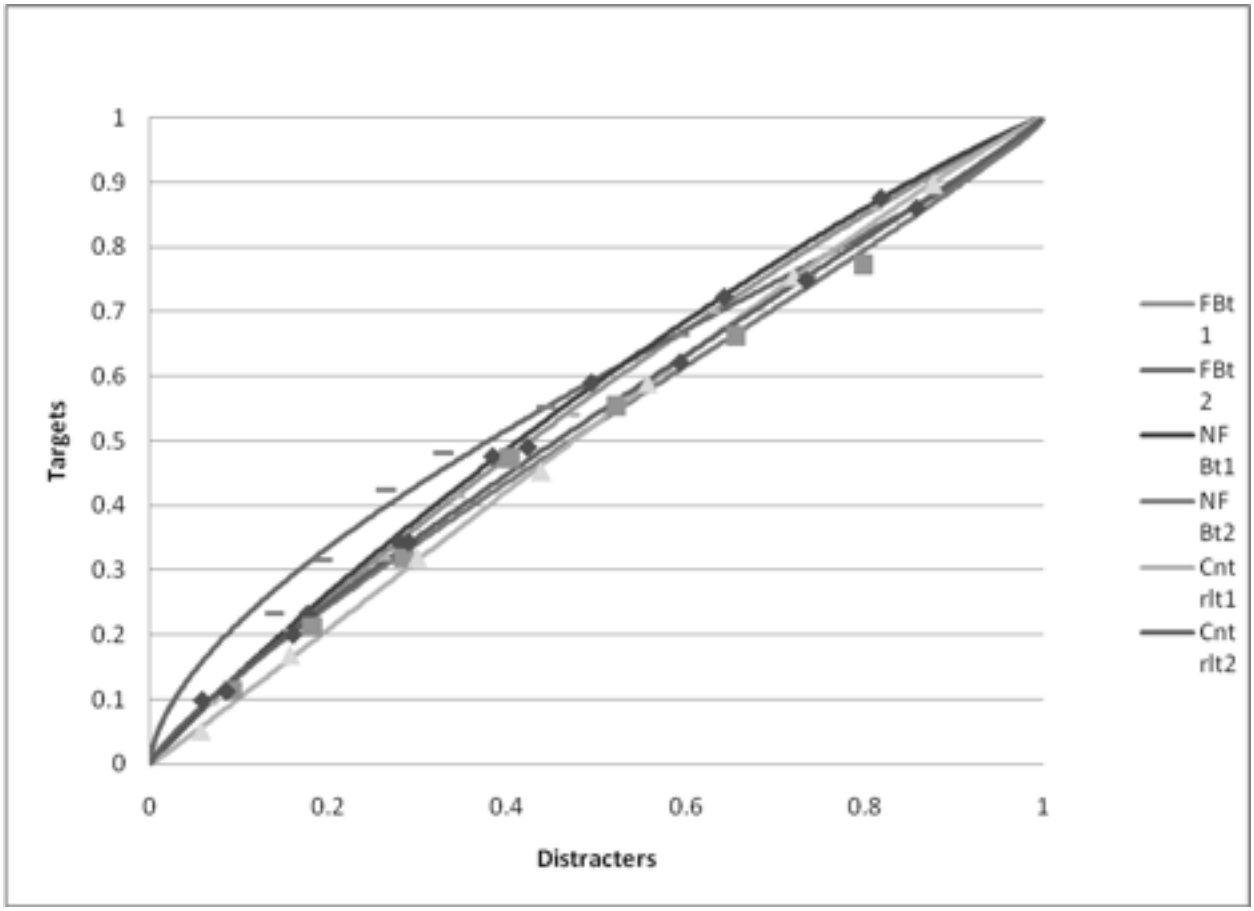

Figure 1. Receiver operator characteristics (ROC) curves for all conditions of Study 1.

Figure 2. Four presentation conditions, clockwise from top left. Head, face, mouth, eyes.

Appendix A

Demographics Questionnaire

What gender are you?
[Male] [Female]

How old are you?
[Open ended response]

What do you consider your racial or ethnic background? (Select all that apply.)
[Asian] [Black] [Hispanic] [Pacific Islander] [White] [Other]

Do you have any children?
[Yes - Biological] [Yes – Non-biological] [No]

Have you served as a jury member?
[Yes] [No]

Have you ever served as a member of law enforcement?
[Yes] [No]

Have you ever interned or worked in the field of social work?
[Yes] [No]

Have you ever worked caring for children in a preschool or elementary school setting?
[Yes] [No]

Do you have any experience as a victim of a non-sexual crime?
[Yes] [No]

Have you had a close friend or family member who was a victim of a non-sexual crime?
[Yes] [No]

Have you ever been in prison?
[Yes] [No]

Each week how many episodes do you watch of “To Catch a Predator”, “Law and Order: SVU”, and other shows that include portrayals of sexual offenders?
[Open ended response]

Do you have any experience as a victim of a sexual crime?

[Yes] [No]

Have you had a close friend or family member who was a victim of a sexual crime?
[Yes] [No]

Did you recognize any of the men in the photographs you just saw?
[Yes] [No]

To what extent do you believe that you can predict if someone is a sexual offender?
[Not at all] – [Totally]

What physical characteristics, social behaviors, or mannerisms do you believe could be used to predict sexual offenders?
[Open ended response]

Appendix B

Brief Mood Introspection Scale

INSTRUCTIONS: Circle the response on the scale below that indicates how well each adjective or phrase describes your present mood.

(definitely do not feel) (do not feel) (slightly feel) (definitely feel)

	XX	X	V	VV		XX	X	V	VV
Lively	XX	X	V	VV	Drowsy	XX	X	V	VV
Happy	XX	X	V	VV	Grouchy	XX	X	V	VV
Sad	XX	X	V	VV	Peppy	XX	X	V	VV
Tired	XX	X	V	VV	Nervous	XX	X	V	VV
Caring	XX	X	V	VV	Calm	XX	X	V	VV
Content	XX	X	V	VV	Loving	XX	X	V	VV
Gloomy	XX	X	V	VV	Fed up	XX	X	V	VV
Jittery	XX	X	V	VV	Active	XX	X	V	VV

Overall, my mood is:

Very Unpleasant — Very Pleasant

-10 –9 –8 –7 –6 –5 –4 –3 –2 –1 0 1 2 3 4 5 6 7 8 9 10