Journal of Social, Evolutionary, and Cultural Psychology
www.jsecjournal.org - 2011, 5(1), 66-91.

Original Article

THE ACCURACY OF INFERENCES ABOUT CRIMINALITY BASED ON FACIAL APPEARANCE

Jeffrey M. Valla
Department of Human Development, Cornell University

Stephen J. Ceci
Department of Human Development, Cornell University

Wendy M. Williams
Department of Human Development, Cornell University

Abstract
A growing body of evidence suggests that rapid, yet accurate, dispositional inferences can be made after minimal exposure to the physical appearance of others. In this study, we explore the accuracy of inferences regarding criminality made after brief exposure to static images of convicted criminals' and non-criminals' faces. We begin with a background of research and theory on the curiously recurrent, and historically controversial, topic of appearance-based inferences of criminality, and a brief justification of our re-opening of the debate about the accuracy of appearance-based criminality judgments. We then report two experiments in which participants, given a set of headshots of criminals and non-criminals, were able to reliably distinguish between these two groups, after controlling for the gender, race, age, attractiveness, and emotional displays, as well as any potential clues of picture origin. Empirical and theoretical implications, limitations, and further questions are discussed in light of these findings.

Keywords: Face processing, criminality judgments, physiognomy, Social Darwinism, born criminal, facial appearance, social inference

AUTHOR NOTE: Please direct correspondence to Jeffrey M. Valla, Department of Human Development, Cornell University, Ithaca, NY 14853, (607)255-9856. Email: jmv34@cornell.edu

©2011 Journal of Social, Evolutionary, and Cultural Psychology

> "*It is possible to infer character from features, if it is granted that the body and the soul are changed together by the natural affections: I say 'natural', for though perhaps by learning music a man has made some change in his soul, this is not one of those affections which are natural to us; rather I refer to passions and desires when I speak of natural emotions. If then this were granted and also that for each change there is a corresponding sign…we shall be able to infer character from features*." (Aristotle, *Prior Analytics*, Translated by A.J. Jenkinson)

The accurate assessment and prediction of personality traits, dispositions, and future behaviors (or "character," as Aristotle refers to them collectively) on the basis of physical appearance has a long and sordid history in philosophical, anthropological, and psychological thought. "Physiognomy," the (pseudo)science of character inferences based on physical appearance, namely facial features, resulted from these early musings by Aristotle and his fellow Greek philosophers. By the mid- to late 19th century interest in the topic began to wane as the scientific credibility of physiognomy was brought into question, a result of the excesses of so-called criminal anthropologists such as Cesare Lombroso (for historical review, see Gould, 1981, pp. 113-145). Shortly thereafter, *The Expression of the Emotions in Man and Animals*, Darwin's (1872) final work, revived interest in the topic by noting the adaptive significance of using physical cues as reliable signals during social interaction. As is well known, Darwin's ideas were subsequently misinterpreted and misused in the Social Darwinism and eugenics movements of the late 19th and early 20th centuries. In the decades following the decline of these movements, evolutionary accounts of nonverbal communication became associated with the wrongs of eugenics, Social Darwinism, restrictive U.S. immigration policies, and the Holocaust, and the idea that humans outwardly reveal reliable signals about their inner selves once again fell out of scientific favor. With the rise of scientific psychology in the 20th century, however, came yet another renewed inquiry into the idea.

"Just why the physical characteristics of individuals should exert so profound an influence over their associates furnishes an interesting topic of speculation," noted Perrin (1921), perhaps highlighting the reason why the topic has continually resurfaced; if others' physical appearances tell us nothing about them, then why do we find them so essential to our social judgments? Asch (1946) not only agreed with Perrin's sentiment about the weight of these inferences on our judgments, but also made the leap to suggest that these judgments are *accurate*: "This remarkable capacity we possess to understand something of the character of another person, to form a conception of him as a human being…with particular characteristics forming a distinct individuality is a precondition of social life." Early research on character inferences gleaned from physical characteristics was promising, as numerous studies found strong relationships between self-ratings and strangers' ratings of personality based on brief, zero-acquaintance interactions (e.g., Cleeton & Knight, 1924; Hunt & Lin, 1967; Norman & Goldberg, 1966; Passini & Norman, 1966).

With the switch in focus to internal processes supplied by the cognitive revolution, research on the topic of appearance-based inferences in the 1970's and early 1980's was concerned mainly with studying appearance-based inferences as forms of stereotyping and their (mostly negative) consequences on behavior in both the perceiver and the perceived (e.g., Aronovich, 1976; Goldstein, Chance, & Gilbert, 1984; Lown, 1977; Mueller, Thompson, & Vogel, 1988; Shoemaker & South, 1978; Shoemaker,

South, & Lowe, 1973). The relative unpopularity of focusing on the *accuracy* of inferences, even with more advanced methodologies available, was possibly due, at least in part, to the societical *Zeitgeist* persisting in the years following the 1960's civil rights movement, as the populace was encouraged to judge books not by their covers. Whatever the reason, questions regarding the accuracy or validity of appearance-based inferences once again fell out of favor, and research on appearance-based inferences was consigend to unconscious negative stereotypes. (To be sure, a small minority of researchers did persist in studying the potential accuracy of rapid social inferences (e.g., Agnew, 1984; Cavior & Howard, 1973), despite the fact that activity of this type of research greatly diminished.) Researchers in the former camp were often explicit in wanting to distance themselves from the latter, often making their audiences aware of their decisions to exclude analyses of accuracy (e.g., Goldstein et al., 1984; Shoemaker & South, 1978; Shoemaker et al., 1973; Yarmey, 1993). Thus, an explicit argument was broadly accepted that social judgements made on the basis of brief exposure, the basis of unfair societal stereotypes and predjudices, were unreliable.

Recent Lines of Research on the Accuracy of Appearance-Based Inferences

The renewed focus on ecological and adaptive approaches in the mid 1980's and early 1990's, especially in the areas of person perception and impression formation (e.g., Baron & Boudreau, 1987; McArthur & Baron, 1983; Zebrowitz & Collins, 1997), reiterated Darwin's prescient hypothesis of an adaptive significance to making quick, yet accurate, assessments of others' traits. Since then research on accuracy has blossomed. Recent research has confirmed the early intuitions of Darwin, Perrin, and Asch, finding that we are able to make surprisingly accurate judgments of others' personalities, behaviors, sexual orientations, and competencies based on minimal interactions or even mere glimpses of them, and that these assessments occur rapidly and automatically (Albright, Kenny, & Malloy, 1988; Ambady & Rosenthal, 1992, 1993; Ambady, Hallahan, & Conner, 1999; Bond, Berry, & Omar, 1994; Kenny, Horner, Kashy, & Chu, 1992; Watson, 1989). Noteworthy is the shift that occurred from regarding rapid character judgements as the result of unfair stereotypes that led to inaccurate social inferences, to regarding them as evidence of ecologically adaptive mechanisms, a slight reversal of the aforementioned trend that followed the cognitive revolution.

While many lines of research on rapid, accurate social cognitions have seemed to escape the longstanding stigma associated with eugenics and social Darwinism discussed above – the "Thin Slice" research paradigm being the prime example – the accuracy of appearance-based inferences still carries the legacy of this stigma, as it continues to be grouped with other archaic morphological approaches, such as physiognomy, craniometry, and phrenology (Hassin & Trope, 2000). The implicit argument seems to be that it is one thing to study accurate inferences about traits from dynamic snippets of people's behavior (e.g., Ambady & Rosenthal, 1992), however brief these snippets are (sometimes <10 seconds), but it is quite another to study accurate inferences based on appearance alone. As Berscheid (1981) argued, research on appearance-based accuracy might seem more credible if society were "not so enamoured of the idea that because a person's appearance *ought* not to make a difference, it *does* not," blaming the dearth of this research on the "naturalistic fallacy" (that is, confusing how things *are* with how they *ought* to be). Thus, despite the burgeoning field of research on rapid social inferences that has occurred since the 1990s, little research has been conducted on the accuracy of

character judgements based solely on static facial information.

As was the case with appearance-based judgments in general, the idea that truthful inferences can be made from static images is broadly regarded as a non-issue in the sense that they are consigned to biases in social information processing rather than being regarded as accurate processing. For instance, recent findings showing that judgments about political candidates' competencies based on facial appearance alone were highly predictive of election outcomes have been explained solely in terms of how these impressions may lead voters to make poor choices, while completely neglecting any mention of the possibility that these competency inferences may actually be accurate (Todorov, Mandisodza, Goren, & Hall, 2005). Of course, in the context of politicians' highly posed headshots the former explanation is probably more likely, but the latter explanation was not even posed as an alternative causal direction for these correlations. The assumption that these rapid, often unconscious decisions were completely uninformed by valid facial cues suggests the authors believe that a consciously-processed list of candidates' pro and con attributes would lead to a more valid decision. In contrast to Todorov et al., however, Nisbett and Wilson (1977) argued that the processing of social stimuli is a largely automatic and unconscious process, and others have argued that unconscious decisions are sometimes better-informed than conscious ones (e.g., Dijksterhuis, 2004; Dijksterhuis, Bos, Nordgren, & Van Baaren, 2006a,b; Dunning & Stern, 1994). The emphasis by researchers such as Todorov et al. on the stereotype-behavior connection, rather than on the potential accuracy of these inferences and the potential advantage of making decisions based on them, may be a recent echo of the historical tendency of researchers to shy away from the possibility of accurate impressions, perhaps out of concern that it harkens back to the stigmas associated with social Darwinism.

Notwithstanding concerns over stigmas and biases, there is reason to believe that faces, even in the absence of dynamic behavior, may be informative. Faces are unique as far as channels of nonverbal communication; they are the first source of information available to a perceiver, and are continuously available during social interaction (Hassin & Trope, 2000; Kleck & Rubenstein, 1975; Zebrowitz, 1999). Facial appearance may not be as informative as some informational sources, but it is easier to "ferret out" than most (Hatfield & Sprecher, 1986). Moreover, until quite recently in human evolution, facial expressions could not be willfully altered and, compared to emotions that change frequently, facial structure is relatively stable (Hassin and Trope, 2000). Also in support of this adaptive value account are findings that personality assessments based on facial appearance reach high levels of consensus both within cultures and between cultures, lending support to the idea of an evolved, universal, adaptive ability to read personality from facial appearance (McArthur & Berry, 1987). Neurological evidence points towards separate neuroarchitectures for reading emotions in faces, versus for inferring intentional stances from faces (Winston, Strange, O'Doherty, & Dolan, 2002). The former would not be ostensibly helpful for inferring stable traits from faces, but the latter would be.

As far as empirical findings in support of the accuracy of judgments based on facial appearance, the evidence is sparse but generally positive. To date, accurate assessments from facial appearance have been found in the domains of intelligence, willingness to deceive, criminality, sexual orientation, certain personality traits (e.g., social dominance, interpersonal warmth), and aggressive tendencies. For instance, Alley (1988) found that people could assess levels of intelligence at better-than-chance levels based on facial appearance alone (though concerns with their methodology have been

raised because of their photograph selection criteria) (Berry & Finch Wero, 1993).

Bond, Berry, and Omar (1994) found that individuals who were predicted by others, based on photos of their faces, to be deceptive, were more likely to participate in research that was described as requiring deception than individuals who were judged to be more trustworthy. More recently and tellingly, research has shown that even in the absence of cheating/cooperation knowledge related to targets' choices in a prisoner's dilemma game, people are better at recognizing targets who cheated than those who cooperated, even if they have never seen targets' faces before (Yamagishi, Tanida, Mashima, Shimoma, & Kanazawa, 2003). The authors suggest that this may indicate the existence of an evolved module specifically for detecting social exchange cheaters based on appearance.

Some researchers using zero-acquaintance paradigms (in which individuals assess strangers' personalities without interacting with them) have posited that the accuracy found in these zero-acquaintance personality assessments (e.g., Ambady & Rosenthal, 1992, 1993) is indicative of accurate trait inferences based on facial appearance. This is somewhat speculative, however, because even in these zero-acquaintance situations more information than just the face (i.e. clothing, body language, hair style, posture) is available to the perceiver. However, Borkenau and Liebler (1992) added a still-frame condition to their zero-acquaintance paradigm and still found high self-other correlations. Recently, Rule, Macrae and Ambady (2009; Rule & Ambady, 2008; Rule, Ambady, Adams, & Macrae, 2008) demonstrated that very brief exposures (50 ms) of faces were sufficient to categorize sexual orientation above chance (60-70% accuracy), and that such brief exposures primed sexual orientation category activation on a subsequent verbal association task.

Appearance-Based Inferences About Criminality

Among the facial appearance-based inferences that have been studied, criminality holds a special place in the history of this line of work, being one of the driving forces behind the longstanding stigma attached to the field (Gould, 1981). As Social Darwinism, eugenics, and genetic determinism began to gain popularity, perhaps it was only a matter of time before the "born criminal" was hypothesized. After noticing similarities between the skull of an infamous criminal and those of "lower" beings and races, Cesare Lombroso, an Italian physician and criminologist, began to hypothesize about the "born criminal," and the physical indicators of this "subhuman species" (e.g., drooping eyes, large ears, protruding jaw, flat nose) (Gibson, 2002). Lombroso's school of thought, later known as the Italian school of criminology, helped spur some of the more inhumane eugenics interventions, such as the sterilization of criminals, in an effort to reduce the prevalence of criminal "genes" in the greater population.

With the emergence of scientific psychology in the early 20th century, Lombrosian theory was shown to be fallacious. With it, as discussed above, went many lines of research which could be construed, or misconstrued, as genetically deterministic. Since then, few studies have attempted to investigate people's stereotypes and abilities related to judging criminality from facial appearance.

The first of this handful of studies was conducted by Thornton (1939) who randomly selected the case records and photographs of twenty criminals from the Nebraska State Penitentiary, and asked a group of participants to note which of four different crimes each criminal had committed. These judgments were found to be correct

at above chance levels. In a study that more closely resembles Lombrosian theory and physiognomy, Kozeny (1962) selected 730 criminals' photographs, and divided them into sixteen crime categories. He then made composite photographs from these categories and, using physiognomic measurement techniques, found that the categories differed significantly in their physiognomic characters, as Lombroso had predicted.

Bull and Green (1980) attempted a stereotype-relevant replication of Thornton's study, but instead used non-criminals' photographs in an effort to see not if observers could identify criminal types, but rather whether the general public shared common beliefs about what different criminal subtypes look like. They found that for crimes of mugging, violent robbery, company fraud, soliciting, car theft, drug possession, and gross indecency, observers chose a particular face more frequently than others, and that different faces were chosen more frequently for each of these crimes. Although these researchers assumed that this was an invalid inference, given that the faces were not of criminals, this does not obviate the possibility that known criminals' faces emit valid cues to their criminal behavior. Goldstein et al. (1984) performed a similar study, and found that observers consensually selected different faces for criminal and non-criminal categories, as well as for subcategories within the criminal and non-criminal categories (mass murderer, armed robber, rapist, medical doctor, and clergyman). The thrust of the conclusions was that face processing was a form of biased social informaation processing. Again, no analysis was done to assess the accuracy of these subjects' judgments, even though the researchers had the data available to do so.

Yarmey (1993) later replicated and extended Goldstein et al.'s findings. He found that people came to a high level of consensus when asked to categorize photos of non-criminals into criminal and non-criminal categories, and also criminal subtypes. And like others before him, he did not examine the validity of his subjects' classifications. Exactly why Bull and Green, Goldstein et al., Yarmey, and others carried out stereotype analyses but not accuracy analyses, when presumably they had the data and/or methods to do so, is puzzling, and one can only hypothesize that these researchers either assumed accuracy was impossible, or wanted to distance their work from accuracy research which may have been seen as a throwback to the excesses of the Lombroso era.

A separate, related line of inquiry that came from stereotype research tested the hypothesis that, rather than specific criminal physiognomies, the connection between appearance and criminality (or deviancy in general) is due to attractiveness. The idea being that, since studies continued to show that stereotyping worked in favor of attractive people, unattractive people received the flipside of this treatment and, lacking the advantages and options of the attractive set, pursued economic and sexual gains through illegal means (Cavior & Howard, 1973). Tests of this hypothesis were generally positive (Agnew, 1984; Cavior, Hayes, & Cavior, 1974; Gross & Crofton, 1977), supporting the Greek philosopher Sappho's assertion that *"what is beautiful is good."* In an extensive review of the early psycholegal literature, Monahan and Loftus (1982) concluded that the effect of a defendant's attractiveness on jurors' deliberations and/or verdicts was among the most consistent effects in the literature, leading to an "attraction-leniency bias" among jurors. Notwithstanding these demonstrations, significant methodological flaws complicated cuasal claims (e.g., confounding socioeconomic status and attractiveness).

Some still furthered these ideas about the role of attractiveness by testing, for instance, the effect of plastic surgery on inmates' recidivism rates, but with mixed findings (Bull & Rumsey, 1988). However, as Berry and Finch Wero (1993) point out, while attractiveness can be a strong predictor of personality, facial appearance and facial

attractiveness are not the same concept, as facial appearance consists of many dimensions in addition to attractiveness. In support of this, Berry (1991) has shown that two faces judged as equally attractive can still elicit qualitatively different impressions. Hence, while attractiveness could be *a* factor in a potential appearance-criminality relationship, logically it need not be *the* factor.

Re-Visiting Appearance-Based Judgments of Criminality

The ecological advantage of being able to spot criminals and criminal subtypes is seemingly obvious. Psychologists taking ecological and evolutionary stances have studied abilities that would have been advantageous in the Environment of Evolutionary Adaptation (EEA), and it is conceivable that identifying social deviants could be one of these abilities. This is not dissimilar from the aforementioned arguments for an evolved cheater detection mechanism (Yamagishi et al., 2003). If humans evolved the ability to spot more immediate signals of danger and attributes such as fertility and immunity in the EEA (Buss & Schmitt, 1993), then it would be reasonable to hypothesize that some of these inferences may be gleaned from facial appearance, being the enriched source of information that it is. In the research to be reported here, we ask whether it is possible that human beings have evolved a way to detect those who are more likely to commit violent and/or non-violent (i.e. economic) crimes. If so, is it also possible that certain sex-specific evolutionary problems have differentially shaped males' and females' abilities in this realm; for instance, are females better at spotting rapists than males?

Since Kozeny's study 46 years ago, there have been no further empirical tests of whether people can identify criminals and criminal subtypes based on appearance, and relatively few tests of the accuracy of inferences from facial appearance in general, with the exceptions of Bond et al. (1994), Hassin and Trope (2000), and Berry and Finch Wero (1993). With the recent focus on the ecological perspective, previous evidence of accurate appearance-based inferences, and the unanswered questions of researchers such as Yarmey, Bull and Green, and Goldstein et al., we believe the question of criminal and criminal subtype identification is still an open one, and with this in mind we conducted two experiments as a first step toward answering it.

In these experiments, we examined participants' abilities to distinguish between criminals and non-criminals, between violent and non-violent criminals, and between specific criminal subtypes (murderers, rapists, thieves, forgers, drug dealers, arsonists, and assailants), employing a large number of controls. In each experiment, we presented participants with a series of headshots, half of which were convicted criminals, the other half being non-criminals. Participants were asked to assess the likelihood each individual committed a crime, the likelihood that they committed a violent or non-violent crime, and the likelihood they committed a specific crime. Participants filled out a demographic survey in order to examine interactions between the ability to "spot" criminals and participant characteristics.

We hypothesized that participants would rate criminals as more likely to commit a crime even when faces were carefully selected to control for attractiveness and other characteristics. We also wondered whether violent criminals would be rated as more likely to commit a violent crime, due to the ecological advantage of being able to detect those who might cause physically harm vs. those who might materially harm you. Last, we predicted that each specific criminal type would be rated more likely to commit their corresponding crimes than all other crimes. Finally, we anticipated a gender x criminal

interaction, whereby female participants would rate rapists as more likely to be a criminal than males would. Below, in the interest of limited space, we briefly summarize the findings from the first experiment, and report the detailed results of the follow-up study which both replicated and extended the pilot study.

Pilot Study Methods

Participants and Materials

In the pilot study, 44 college students were shown 32 headshots, 16 depicting criminals, and 16 non-criminals. The latter came from the "NimStim" photo catalogue, a collection of headshots depicting a series of emotional states (Tottenham, 2007). We collected all the photos that fit the following criteria: Caucasian, male, between the ages of 20 and 25, no facial scars, tattoos, or other markings, and with little or no facial hair. From the photos matching the above criteria, we randomly selected 16 individuals and used the "neutral" emotion pictures for each to control for emotional expression. The criminal photographs came from the Missouri, Montana, Michigan, and Florida online criminal offender databases (https://web.mo.gov/doc/offSearchWeb/; http://app.mt.gov/conweb/; http://www.state.mi.us/mdoc/asp/otis2.html; http://www.dc.state.fl.us/AppCommon/). We selected murderers, rapists, thieves, and forgers who met all of the above control criteria, as well as having been convicted of only one of these crimes. Photographs were edited to remove background, show only the heads, and maintain a consistent photo quality, and remove differences in lighting, graininess, photo quality, etc.

Procedures

The order of photo presentation was randomized and participants rated each photo from 1 (Extremely Unlikely) to 9 (Extremely Likely) on the likelihood that the depicted person committed each of four crimes (murder, rape, theft, and forgery). Thus, each participant made a total of 128 ratings (4 ratings x 32 photos). Participants were told that some of the depicted individuals were non-criminals, some were criminals, and that the criminals committed only one of the four crimes; no mention was made of what proportion were criminals or each criminal subtype to avoid influencing decision criteria. Participants were asked if they thought it was obvious that any of the pictures were mug shots.

14 independent raters (6 male, 8 female) who did not participate in the criminality rating were asked to rate each of the photos from 1 (extremely unattractive) to 10 (extremely attractive). These ratings (table 1) were averaged by photo, and covaried in analyses.

Pilot Study Results

A series of mixed effects ANOVAs, with Subject entered as a random effect, and Photo Category (Criminal/Non-Criminal) and Attractiveness entered as fixed effects, revealed that criminals were rated as significantly more likely than non-criminals to have committed murder, rape, theft (p's $< .0001$), and forgery ($p = .04$). There were zero interactions between ratings given to each category and yes/no responses to the question,

"Was it obvious that some photos were mug shots?." Thus, there were no differences in the ratings of those claiming to notice extraneous cues.

We also examined participants' ability to distinguish violent from non-violent criminals, using a mixed effects model with Subject entered as a random effect, and Attractiveness and Photo Category (Violent/Non-Violent) as fixed effects. Contrary to the hypothesis, there were no differences between these two groups on ratings of murder--participants did not distinguish violent from non-violent criminals.

Table 1. Summary Statistics of Crime Likelihood ratings, by Photo Category and Rating Type, Pilot Study

		Likelihood of Committing:							
		Murder		Rape		Forgery		Theft	
Photo Category	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Murderers	176	4.57	2.15	5.22	2.1	4.55	2.11	5.56	1.87
Rapists	176	4.24	2.29	4.51	2.27	4.61	2.2	5.45	2.12
Violent Criminals	352	4.4	2.23	4.86	2.21	4.58	2.15	5.5	2
Forgers	176	4.9	2.33	5.47	2.22	4.49	2.05	5.68	1.9
Thieves	176	4.28	2.22	4.66	2.17	5.04	2.2	5.23	2.13
Non-Violent Criminals	352	4.59	2.3	5.07	2.23	4.76	2.14	5.45	2.03
Criminals	704	4.5	2.26	4.96	2.22	4.67	2.15	5.48	2.01
Non-Criminals	704	3.85	2.12	4.38	2.18	4.48	2.05	4.86	2.11
All Photos	1408	4.17	2.22	4.67	2.22	4.58	2.1	5.17	2.08

In addition, we found that murderers were rated just as likely to have committed murder as rapists, forgers, and thieves, and rapist photos were rated significantly *less* likely to have committed rape than were murderers, thieves, and forgers, combined ($p < .001$). Thieves were rated as no more or less likely to have committed theft than murderers, rapists, and forgers combined, and forgers were not rated differently than other criminal types to commit forgery.

Finally, we repeated the above analyses separately for each sex. The only significant effect involved females' rape likelihood ratings. Comparing females' rape ratings for rapists to those given to the other three criminal categories revealed that rapists were rated significantly *less* likely than other criminal types to have committed rape ($p = .04$).

Pilot Study Discussion

Thus, the main finding from this pilot study was that participants rated convicted murderers, rapists, thieves, and forgers as more likely to have committed crimes than non-criminals, and they were able to do so without being given any information other than static cropped images of their faces. These results could not be attributed to differences in attractiveness, race, gender, age, facial hair, hairstyle, or photo quality between these groups. On the other hand, participants lacked the finer-grained ability to

discriminate between violent and non-violent criminals, and between the specific crimes committed by these individuals. Female participants had particular difficulty identifying rapists, who received significantly *lower* rape likelihood ratings than other criminals--no greater than those of non-criminals. That is, females rated rapists as significantly *less* likely to have committed rape than the other criminals. It is possible that rapists' non-threatening appearance is precisely how they are able to gain access to unsuspecting victims.

Main Study

The main study was designed to both replicate and extend the above finding that participants can distinguish between criminals and non-criminals, using an improved methodology. We employed new criminal categories (except rape, which we once again included, given the interesting finding), new criminal faces, and new non-criminal faces, to add generality to the above findings, and modified procedures that allowed us to use signal detection theory to address a question that arose.

Main Study Methods

Participants and Materials

36 participants (33 female, 3 male; ages 19 through 26), were recruited from a psychology class. We selected a different set of 32 headshots (16 criminal, 16 non-criminal) for this experiment. The non-criminal photos were selected from a collection provided by a colleague at another university, rather than from the NimStim catalogue used in the pilot study. This new catalogue was preferable because its photos varied more in photo source than the NimStim. As in the first experiment, we pared down the pool of potential photos to include only individuals who were Caucasian, male, with no facial scars, tattoos, or other markings, and with little or no facial hair, and then randomly selected 16 individuals from this pool. It was necessary to extend the upper age limit from 25 in the first experiment to 29.

The criminal photographs were obtained from the same offender databases in the pilot study. To increase the generalizability, we completed a series of targeted searches for rapists, assailants, arsonists, and drug dealers who met the above control criteria and had been convicted of only one of these crimes. Rapists were included to retest the hypothesis regarding a Gender x Rapist interaction, this time using only rapists who committed a violent form of rape, as our original set of rapists may have contained statutory, or non-violent, rapists.

The non-criminal and criminal photographs were edited with Macromedia Fireworks to ensure that all photos had a consistent background color, showed only the heads of the depicted individuals, and had a consistent photo quality (lighting, graininess, contrast, etc.). These photos were presented randomly in a Powerpoint presentation, one photo per slide.

Procedures

Participants were told that they would see 32 photos, some of which were non-criminals, and some of which had been convicted of one of the four crimes. They were not told what proportion of the photos belonged to each category.

Participants were asked to a) rate the likelihood that each individual had committed *any* crime, from 1 (Extremely Unlikely) to 7 (Extremely Likely); b) *if* they thought the person committed a crime (indicated by a rating of 5 or greater on the first question--the "Likely" half of the Likert scale), they were asked if the individual committed a violent or non-violent crime, and how confident they were from 1 (Completely Unsure) to 4 (Completely Sure); c) *if* they thought the individual committed a *violent* crime, they were asked whether it was rape or assault, and indicate their confidence from 1 to 4; d) *if* they thought it was a *non-violent* crime, they were asked whether it was arson or a drug offense, and rate their confidence. Participants were presented with the 32 criminal and non-criminal photos, one photo at a time, for 20-30s per photo, on a projection screen. They were also asked if it was obvious (yes or no) any of the pictures were mug shots. Finally, participants were re-presented the 32 photos, and asked to rate the attractiveness on a scale from 1 (Extremely Unattractive) to 10 (Extremely Attractive).

An additional 13 participants rated each photo on how happy, sad, angry, surprised, pleasant, and aroused it looked, on scale from 1 (Not At All) to 7 (Extremely). These ratings were averaged by photo and used to control for any affect differences between categories.

Main Study Results

Criminality Likelihood Ratings

A mixed effects model was used to test the effect of photo category on criminality ratings (table 2), with Subject as a random effect, and Photo Category (criminal, non-criminal) as a fixed effect. To rule out any potential influence of affective display differences between the criminals and non-criminals, the averaged affective ratings of the 13 independent raters of happiness, sadness, anger, surprise, pleasantness, and aroused were factor analyzed and the first three principal components (table 3) were entered, along with attractiveness ratings, as fixed effect covariates. The main effect of criminals rated more likely to have committed a crime than non-criminals was significant, $F(1, 1103) = 4.16$, $p = .048$, even after controlling for attractiveness, $F(1, 953) = 106.25$, $p < .0001$, and affect display ($F(1, 1106) = 67.82$, $p<.0001$, $F(1, 1102) = 19.61$, $p<.0001$, and $F(1, 1104) = 25.21$, $p < .0001$, for the first, second, and third principal components, respectively).

Table 2. Summary Statistics of Crime Likelihood ratings, by Photo Category, Main Study

Photo Category	*N*	*M*	*SD*
All Photos	1144	4.29	1.28
Criminals	572	4.55	1.25
Non-Criminals	572	4.03	1.25
Non-Violent Criminals	286	4.72	1.28
Violent Criminals	286	4.38	1.20
Arsonists	144	4.36	1.43
Assailants	142	4.82	1.09
Drug Dealers	142	5.08	0.98
Rapists	144	3.95	1.15

Violent Versus Non-Violent Criminals

A mixed effects model testing the effect of violent/non-violent category on criminality ratings, controlling for attractiveness and affect display, showed non-violent criminals were rated *more* likely to commit a crime than violent criminals $F(1, 534) = 5.66$, $p = .013$.

A mixed effects model comparing the ratings of the four criminal categories, again controlling for subject, attractiveness and affect display differences, found a significant effect of individual photo types, $F(1, 1102) = 8.59$, $p < .0001$ on criminality ratings. A Tukey-Kramer honestly significant difference test revealed the only significant differences among the five photo categories were the greater ratings given to convicted drug dealers in comparison to non-criminals and rapists, with none of the other contrasts being reliable. Of all photo categories, including non-criminals, rapists were rated as *least* likely to commit a crime, but not significantly so.

Signal Detection Analyses

The "Likely to be a Criminal" and "Unlikely to be a Criminal" halves of the Likert scale were recoded as hits, misses, false alarms, and correct rejections. Ratings of 4 (Neutral/No Opinion) were coded as neither hits nor misses. Hits, misses, false alarms, and correct rejections were used to measure accuracy in terms of discriminitive sensitivity, or *d*' (*z* (Hits/Hits + Misses) – *z* (False Alarms/False Alarms + Correct Rejections) separate from response biases (□□□□ (Participants who are biased toward claiming that every photo depicts a criminal would exhibit a high number of hits but also a high number of false alarms and their □ would be low.) Participants were slightly sensitive (*d*' = 0.5) to detecting criminals.

Table 4. Summary of Criminal/Non-Criminal Hits, Misses, False Alarms, and Correct Rejections, by Photo Category, Main Study

Photo Category	Hits	Misses	False Alarms	Correct Rejections
Criminals	368	142	N/A	N/A
Non-Criminals	N/A	N/A	246	214
Non-Violent Criminals	198	59	N/A	N/A
Violent Criminals	170	83	N/A	N/A
Arsonists	80	46	N/A	N/A
Assailants	106	21	N/A	N/A
Drug Dealers	118	13	N/A	N/A
Rapists	64	62	N/A	N/A

A violent criminal discriminability index (*d'*) was calculated for correctly identified criminals (Table 5). "False alarms" from miscategorized non-criminal photos were excluded from these analyses. The question addressed was, "Among criminal photos correctly identified as criminals, could participants tell the difference between violent and non-violent criminals?" Participants were equivocally sensitive, overall, to violent/non-violent distinctions, $d' = 0.04$.

Table 5. Summary of Violent/Non-Violent Hits, Misses, False Alarms, and Correct Rejections, by Photo Category, Main Study

Photo Category	Hits	Misses	False Alarms	Correct Rejections
All Criminals	107	146	176	81
Non-Violent Criminals	N/A	N/A	176	81
Violent Criminals	107	146	N/A	N/A
Arsonists	N/A	N/A	86	40
Assailants	72	55	N/A	N/A
Drug Dealers	N/A	N/A	90	41
Rapists	35	91	N/A	N/A

For analyses of participants' accuracy identifying individual criminal types among criminals correctly identified as criminals data are shown in Table 6. This analysis revealed little sensitivity in distinguishing individual criminal types, $d' = 0.18$.

Table 6. Summary of Individual Crime Hits, Misses, False Alarms, and Correct Rejections, by Photo Category, Main Study

Photo Category	Hits	Misses	False Alarms	Correct Rejections
All Criminals	110	265	265	860
Arsonists	20	62	36	257
Assailants	46	63	83	183
Drug Dealers	32	86	63	194
Rapists	12	54	83	226

Separately by Gender

These models were repeated separately for males and females. There was no rating difference between criminal and non-criminals, and violent and non-violent categories in both males and females, but in females the higher ratings given to criminals over non-criminals approached significance, $F(1, 1010) = 3.56$, $p = .06$. On the other hand, signal detection analyses showed that males' discriminability was slightly better (d'=0.55) than females' (d'=0.49). In males, there were no differences in criminality ratings between crime categories, but in females a significant difference did arise, and linear contrasts showed that females rated rapists as significantly *less* likely to have committed a crime than the other three criminal types, $F(1, 1007) = 21.19$, $p < .0001$, and significantly less likely to have committed a crime than non-criminals, $F(1, 1008) = 3.97$, $p = .04$.

Although the above analyses ruled out the possibility that the use of extraneous cues could explain the criminal rating differences found between criminal and non-criminal photos, we were interested in group differences between participants who answered "yes" (N=24), and "no" (N=11) to the question regarding "knowing" that some photos were mug shots.

A mixed effects model testing for a Knowledge of Mugshot x Photo Category interaction found that it was not significant, $F(1, 1070) = 1.85$, $p = .17$, and, interestingly, the direction indicated that participants who said that they *could* tell which photos were mug shots rated criminal photographs as *less* likely to have committed a crime. This is further supported by signal detection analyses done separately for these two groups; individuals who professed knowledge of picture origin had sensitivity (d') of 0.5, and those who did not had a sensitivity of 0.56, meaning those who could not distinguish between picture quality of mugshots and non-mugshots were more sensitive identifying criminals.

Main Study Discussion

The main study replicated and added to the main finding from our initial, pilot experiment: participants were able to distinguish between criminals and non-criminals based on static facial images alone, controlling for race, age, facial hair and markings, attractiveness, or any other obvious indicators of criminal status. This finding was replicated with a different set of photos, a refined rating system, affect control measures, and using three new criminal types. Thus, while replicating the main finding from the

initial experiment, we also ruled out factors that may have affected the ratings in the first experiment, namely emotional display cues that may have been specific to the situation in which criminals were photographed. We also showed that our previous finding was not a function of the particular photos or criminal types used in the first experiment.

By contrast, signal detection sensitivity in distinguishing violent and non-violent criminals was weak to nonexistent, and participants believed that non-violent criminals were more likely to have committed a crime than violent criminals. This was in contrast to the first experiment in which participants believed that violent and non-violent criminals were equally likely to have committed rape, murder, forgery, and theft. There was little support for participants' ability to identify individual criminal subtypes, with near-zero sensitivity in discriminating individual criminal types. While this lack of a more fine-grained detection ability ran contrary to our hypotheses, our findings do lend some support to Gottfredson and Hirschi's "general theory of crime" (Gottfredson & Hirschi, 1990; Hirschi & Gottfredson, 1994). The basic idea underlying this theory is that the propensity for an individual to commit crime is generalizable to most if not all other types of crimes, which helps explain why those who commit one crime are also more likely to commit other crimes; behavioral self-control (or lack thereof) is, in turn, said to underlie this general propensity to commit crimes of all types. Importantly, in light of our evaluation of different appearance-behavior models, Gottfredson and Hirschi ascribe at least part of self-control to early learning experiences. In other words, and in contrast to the "born criminal," genetic predeterminism model, the environment plays a major part in shaping self-control, and thus criminal tendencies.

As in the initial study, participants had special difficulty identifying rapists, but whereas the pilot experiment found a difficulty in identifying rapists as rapists, the second experiment extended this difficulty to identifying them as criminals in general. In fact, participants rated rapists as the least likely of all photo types, including non-criminals, to have committed *any* crime.

Interestingly, when analyzed separately for each sex, this particular difficulty in identifying rapists remained for females, but not males. From an ecological perspective, one could hypothesize that rapists who successfully gain access to their eventual victims are able to do so partly by appearing non-threatening. In fact, if our finding of a sex difference in ability to spot rapists has any ecological significance, rapists may appear non-threatening in ways that are particularly deceptive to females. One possibility, based on our data, is attractiveness: rapists in our main experiment were rated, on average, more attractive than all other photo categories. In other words, successful male rapists may gain access to their victims because females find them non-threatening due to their attractiveness. This is purely speculative, however, as the rapists in our study may not be representative of the attractiveness of all rapists. Still, it raises an interesting possibility for further inquiry.

Last, we found that, ironically, those participants who claimed to be able to tell which photos were mug shots actually rated criminal photos as *less* likely to have committed a crime, and showed *lower* discriminability sensitivity than those who claimed to be blind to picture source. This finding strengthens our belief that participants did not distinguish criminals from non-criminals by relying on extraneous cues.

General Discussion

Across two different sets of criminal/non-criminal stimuli, constituting seven different criminal subtypes, different measurement techniques, and controlling for myriad physical and affective cues, we found that participants presented with brief exposures to static cropped facial images of convicted criminals and non-criminals were able to reliably distinguish between them. However, participants were unable to accurately judge more nuanced distinctions of criminality, namely the violent/non-violent status, and the crime committed.

Outstanding Issues Regarding Stimuli

Although we went to great lengths in constructing a stimulus set and controlled for facial covariates of criminality and beliefs that photos were mug shots, we realize that skepticism toward our stimulus set will understandably remain for some readers, as it is possible to argue that there remained some attribute that raters were able to detect. If true, then perhaps the results of these experiments were not proof that criminal faces were unlike noncriminal faces.

Two findings militate against this interpretation: first, raters were unable to identify which photos were in fact mugshots, and not only was there no significant association between their choices and their judgments, but in absolute terms those professing mugshot knowledge also rated criminal photos as less "criminal-like" than non-criminal photos, unlike those claiming mugshot knowledge ignorance. Second, if raters were somehow able to detect characteristics of mugshots, and base their criminality judgments on them, this would be incompatible with the repeated finding that female raters were reliably *less* likely to identify rapists. It would be implausible to posit that raters in these experiments somehow detected criminal mugshots for all crimes except rape because the latter were somehow physically different from the usual mugshots in physical attributes.

In addition, photos were not selected randomly from offender databases. Rather, we selected photos that fit our criteria of Caucasian, Age 20-28, no facial hair or tattoos, average attractiveness, no menacing expression, etc. Restricting photos to these criteria, excluded the vast majority of criminals who "looked criminal." Thus, we deliberately stacked the deck against our hypotheses by selecting convicts who didn't look menacing, "seedy," or deceitful. It is fairly rare that a male, between the ages of 20 and 25, Caucasian, with no facial hair or tattoos, and of average attractiveness commits murder. Evidence that we stacked the deck against our hypotheses by choosing the least criminal-looking criminals comes from a second replication we did following this main study. We repeated the procedures described above but used randomly selected criminal faces that met all of the criteria from the same offender databases but were not screened for affective expressions. These randomly selected photos were rated as much more likely to have committed a crime (M = 4.87, SD = 0.68) than the criminal photos used in the earlier studies (M = 3.75, SD = 0.49), t (27.23) = 5.38, p<.0001, the effect size being enormous, d = 1.9. In assembling photos some valid sources of variance were expunged, rendering the present findings even more impressive.

One could also argue that perhaps participants were simply perpetuating the effects of criminal stereotyping that original jurors employed to convict innocent defendants, thereby identifying innocent individuals as convicts who were wrongly

convicted in the first place. While possible, the wrongful conviction rate is estimated to be less than 5% (Gross & O'Brien, 2007), thus it is unlikely that inaccurate stereotyping by original jurors led to these men's convictions and it was perpetuated by our raters.

Third, it is possible that aspects of picture quality were subconsciously processed by individuals who professed blindness with respect to picture origin. This argument rests upon the assumption that photo quality varied more between criminals and non-criminals than within each category. However, mug shots came from precincts around the country, using different cameras and taken with different lighting; and non-criminal photos were taken with cameras of different quality and with different lighting.

We have included the stimulus set from the main study (Appendix A), presented without photo category identification, and invite readers to try to identify which photos are mug shots and which are non-criminals. The key is provided at the end.

Acknowledgements

We would like to thank Matthew Belmonte, John Cleese, and Melissa Ferguson for feedback on a draft of this paper. We thank Nim Tottenham at the University of Minnesota, Twin Cities, for providing us with the NimStim Face Stimulus Set. The development of the NimStim Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the stimulus set.

Received May 10, 2010; Revision received January 20, 2011; Accepted January 25, 2011

References

Agnew, R. (1984). Appearance and delinquency. *Criminology: An Interdisciplinary Journal, 22,* 421-440.

Albright, L., Kenny, D., & Malloy, T. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 55*, 387-395.

Alley, T. R. (1988). Physiognomy and social perception. In T. R. Alley (Ed.) *Social and Applied Aspects of Perceiving Faces* (pp. 167-186). Hillsdale, NJ: Lawrence Erlbaum.

Ambady, N., Hallahan, M., and Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69,* 518-529.

Ambady, N., & Rosenthal, R. (1992). Thin slices of behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 2,* 256-274.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64,* 431-441.

Aronovich, C. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *Journal of Social Psychology, 99,* 207-220.

Baron, R., & Boudreau, L. (1987). An ecological perspective on integrating personality and social psychology. *Journal of Personality and Social Psychology, 53*, 1222-1228.

Berry, D., & Finch Wero, J. (1993). Accuracy in face perception: A view from ecological psychology. *Journal of Personality, 61,* 504-518.

Berscheid, E. (1981). An overview of the psychological effects of physical attractiveness. In G. Lucker, K. Ribbens, & J. McNamara (Eds.), *Psychological aspects of facial form*. Ann Arbor: University of Michigan Press.

Bond, C., Berry, D., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology, 15*, 523-534.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*, 645-657.

Bull, R., & Green, J. (1980). The relationship between physical appearance and criminality. *Medicine, Science, and the Law, 20,* 79-83.

Bull, R., & Ramsey, N. (1988). *The Social Psychology of Facial Appearance,* New York: Springer-Verlag.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: A contextual evolutionary analysis of human mating. *Psychological Review, 100,* 204-232.

Cavior, N., & Howard, L. (1973). Facial attractiveness and juvenile delinquency. *Journal of Abnormal Child Psychology, 1,* 202-213.

Cleeton, G., & Knight, F. (1924). Validity of character judgments based on external criteria. *Journal of Applied Psychology, 8,* 215-231.

Colvin, C. (1993). "Judgeable" people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology, 64,* 861-873.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*, London: John Murray.

Dijksterhuis, A., Bos, M., Nordgren, L., & Van Baaren, R. (2006a). On making the right choice: The deliberation-without-attention effect. *Science, 311,* 1005-1007.

Dijksterhuis, A. & Nordgren, L.F. (2006b). A theory of unconscious thought. *Perspectives on Psychological Science, 1*, 95-109.

Dijksterhuis, A. (2004). Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality Social Psychology, 87*, 586- 598.

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology, 67*, 818-835.

Gibson, M. (2002). *Born to Crime: Cesare Lombroso and the Origins of Biological Criminology*. Westport, CT: Praeger

Goldstein, A., Chance, J., & Gilbert, B. (1984). Facial stereotypes of good guys and bad guys: A replication and extension. *Bulletin of the Psychonomic Society, 22*, 549-552.

Gottfredson, M., & Hirschi, T. (1990). *A General Theory of Crime*. Stanford: Stanford University Press.

Gould, S. J. (1981). *The Mismeasure of Man*. NY: Norton.

Gross, S., & O'Brien, B. (2007). Frequency and predictors of false conviction: Why we know so little, and new data on capital cases. 2nd Annual Conference on Empirical Legal Studies Paper; U of Michigan Public Law Working Paper No. 93; MSU Legal Studies Research Paper No. 05-14.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78,* 837-852.

Hatfield, E., & Sprecher, S. (1986). *Mirror, mirror…: The importance of looks in everyday life*. Albany: State University of New York.

Hirschi, T., & Gottfredson, M. (1994). *The Generality of Deviance*. New Brunswick: Transaction.

Hunt, R., & Lin, T. (1967). Accuracy of judgments of personal attributes from speech. *Journal of Personality and Social Psychology, 6*, 450-453.

Kagan, J. (1989). *Unstable Ideas: Temperament, Cognition, and the Self*. Cambridge, MA: Harvard University Press

Kenny, D., Horner, C., Kashy, D., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues and stability. *Journal of Personality and Social Psychology, 62,* 88-97.

Kleck, R., & Rubenstein, C. (1975). Physical attractiveness, perceived attitude similarity, and interpersonal attraction in an opposite-sex encounter. *Journal of Personality and Social Psychology, 31,* 107-114.

Lown, C. (1977). Legal approaches to juror stereotyping by physical characteristics. *Law and Human Behavior,* 1, 87-100.

Malatesta, C., Fiore, M., & Messina, J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging, 1,* 64-69.

Manning, J. T., Baron-Cohen, S., Wheelwright, S., & Sanders, G. (2001). The 2nd to 4th digit ratio and autism. *Developmental Medicine and Child Neurology, 43,* 160-164.

McArthur, L., & Baron, R. (1983). Toward an ecological theory of social perception. *Psychological Review, 90*, 215-238.

McArthur, L., & Berry, D. (1987). Cross-cultural agreement in perceptions of babyfaced adults. *Journal of Cross-Cultural Psychology, 18*, 165-192.

McFadden, D., Loehlin, J. C., Breedlove, S. M., Lippa, R. A., Manning, J. T., & Rahman, Q. (2005). A reanalysis of five studies on sexual orientation and the relative length of the 2nd and 4th fingers (the 2D:4D ratio). *Archives of Sexual Behavior, 34*, 341–356.

Monahan, J., & Loftus, E.F. (1982). The Pyschology of Law. *Annual Review of Psychology*, *33*, 441-475.

Mueller, J., Thompson, W., & Vogel, J. (1988). Perceived honesty and face memory. *Personality and Social Psychology Bulletin, 14*, 114-124.

Muller, U., & Mazur, A. (1997). Facial dominance in homo sapiens as honest signaling of male quality. *Behavioral Ecology*, *8*, 569-579.

Nisbett, R., & Wilson, T. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35,* 250-256.

Norman, W., & Goldberg, L. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4,* 681-691.

Passini, F., & Norman, W. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology, 4,* 44-49.

Rosenthal, R. (Ed.). (1979). *Skill in Nonverbal Communication: Individual Differences*. Cambridge, MA. Oelgeschlager, Gunn, & Hain.

Rule, N.O., Ambady, N., Adams, R.B., Jr., & 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. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology, 44*, 1100–1105.

Rule, N.O, Macrae, C.N., & Ambady, N. (2009). Ambiguous group membership is extracted automatically from faces. *Psychological Science, 20*(4), 441-443.

Sanders, G., Bereczkei, T., Csatho, A., & Manning, J. (2005). The ratio of the 2nd to 4th finger length predicts spatial ability in men but not women. *Cortex, 41*, 789-795.

Shevlin, M., Walker, S., Davies, M., Banyard, P., & Lewis, C. Can you judge a book by its cover? Evidence of self-stranger agreement on personality at zero acquaintance. (2003). *Personality and Individual Differences, 35,* 1373-1383.

Shoemaker, D., & South, D. (1978). Nonverbal images of criminality and deviance: Existence and consequence. *Criminal Justice Review, 3*, 65-80.

Shoemaker, D., South, D., & Lowe, J. (1973). Facial stereotypes of deviants and judgments of guilt and innocence. *Social Forces, 51,* 427-433.

Thornhill, R. and S. W. Gangestad. (1999). The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and Human Behavior, 20,* 175-201.

Thornton, G. (1939). The ability to judge crimes from photographs of criminals. *Journal of Abnormal and Social Psychology, 34,* 378-383.

Todorov, A., Mandisodza, A., Goren, A., & Hall, C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308,* 1623-1626.

Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology, 57,* 120-128.

Winston, J., Strange, B., O'Doherty, J., & Dolan, R. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature and Neuroscience, 5,* 277-295.

Yamagishi, T., Shigehito, T., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: evidence that cheaters may look different from cooperators. *Evolution and Human Behavior. 24,* 290-301.

Yarmey, A. (1993). Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Applied Cognitive Psychology, 7,* 419-431.

Zebrowitz, L. (1999). *Reading Faces: Window to the Soul?* Boulder: Westview

Zebrowitz, L., & Collins, M. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Bulletin, 3,* 204- 223.

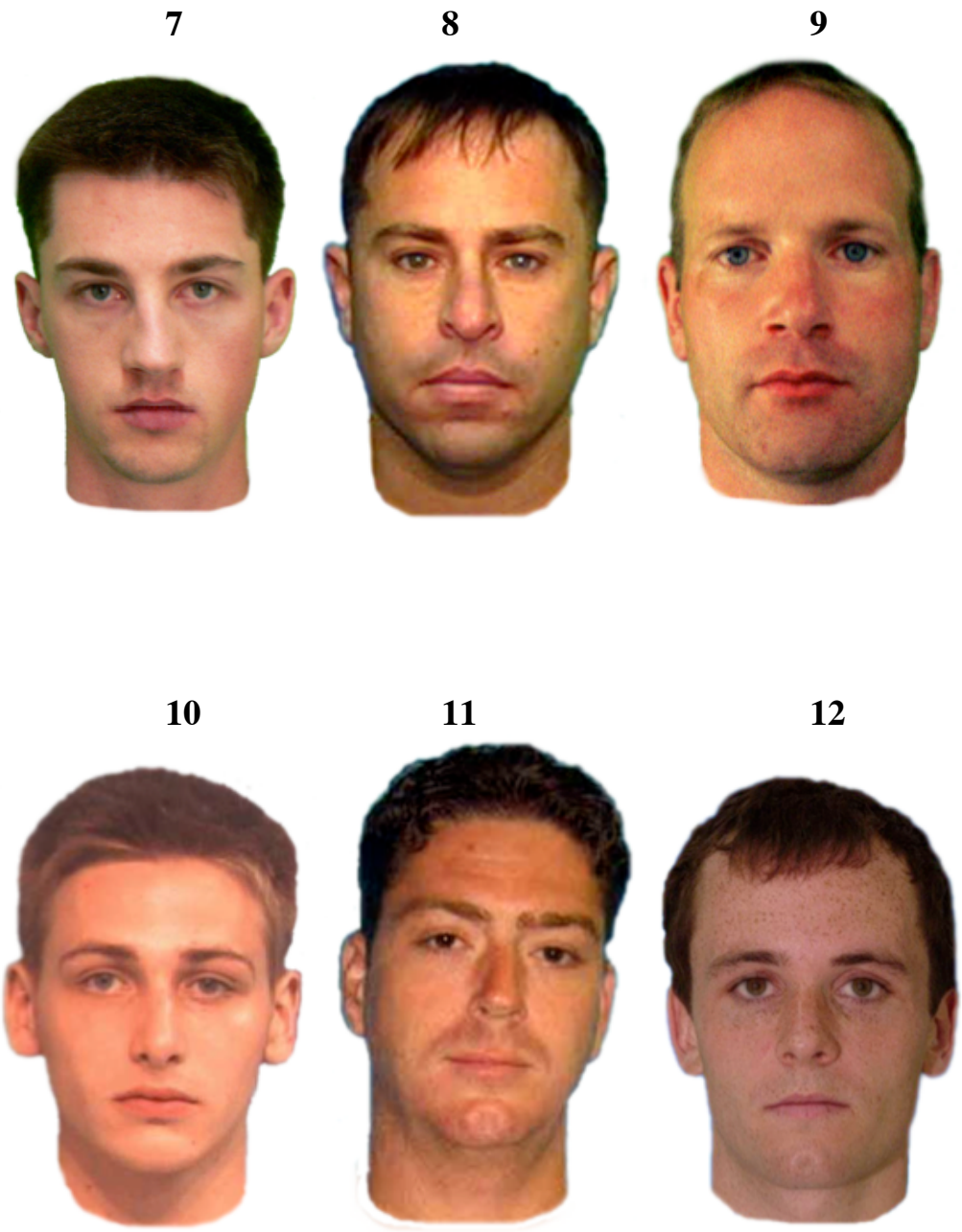
Zebrowitz, L., Collins, M., & Dutta, R. (1998). The relationship between appearance and personality across the life span. *Personality and Social Psychology Bulletin, 24,* 736-749.

Appendix

Appendix A. Stimulus Set from Main Study (Key Provided at End)

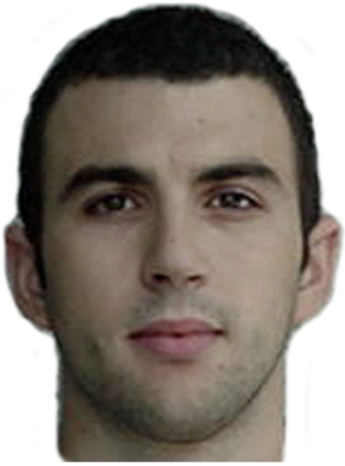
1 **2** **3**

4 **5** **6**


7
8
9
10
11
12

13

14

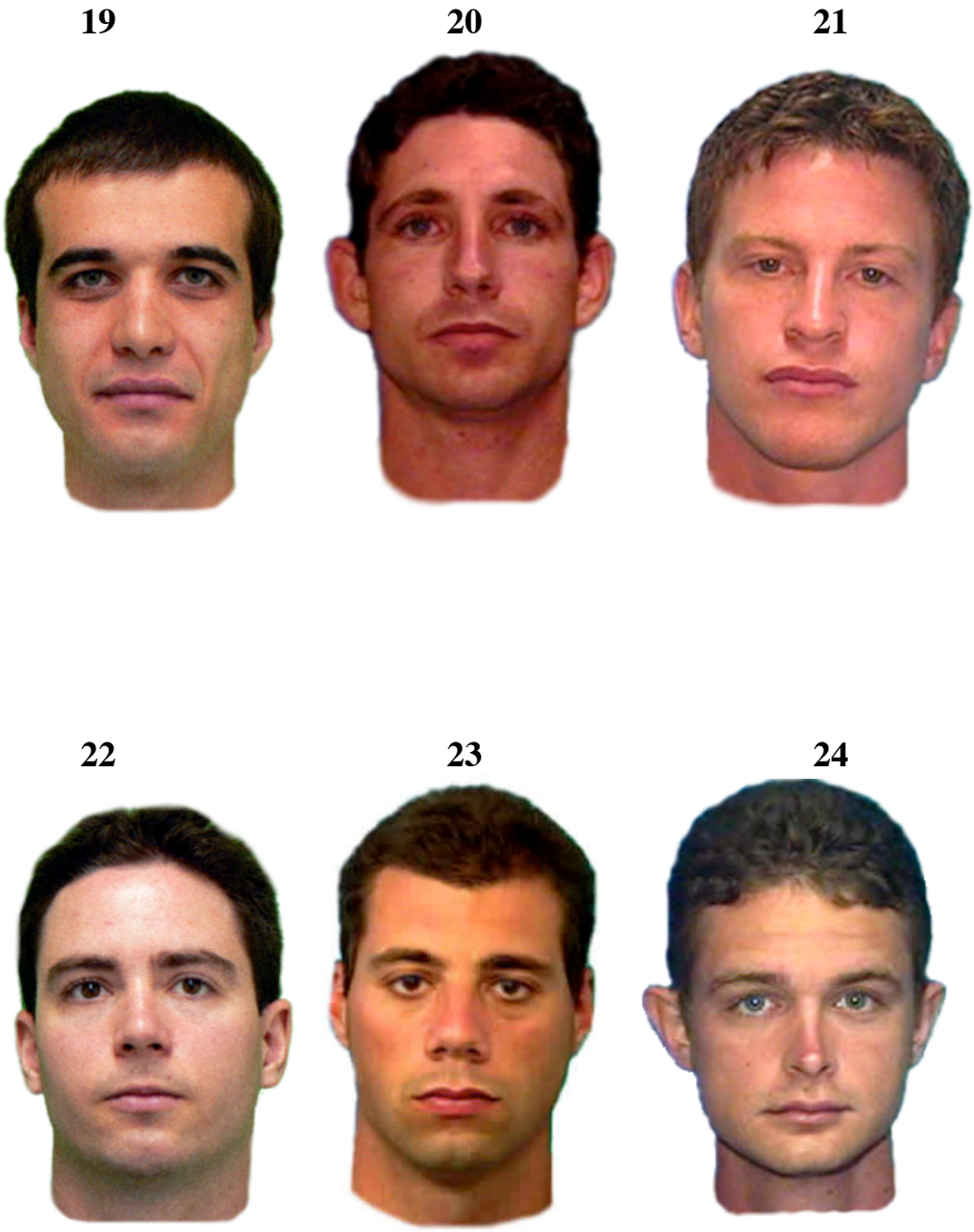


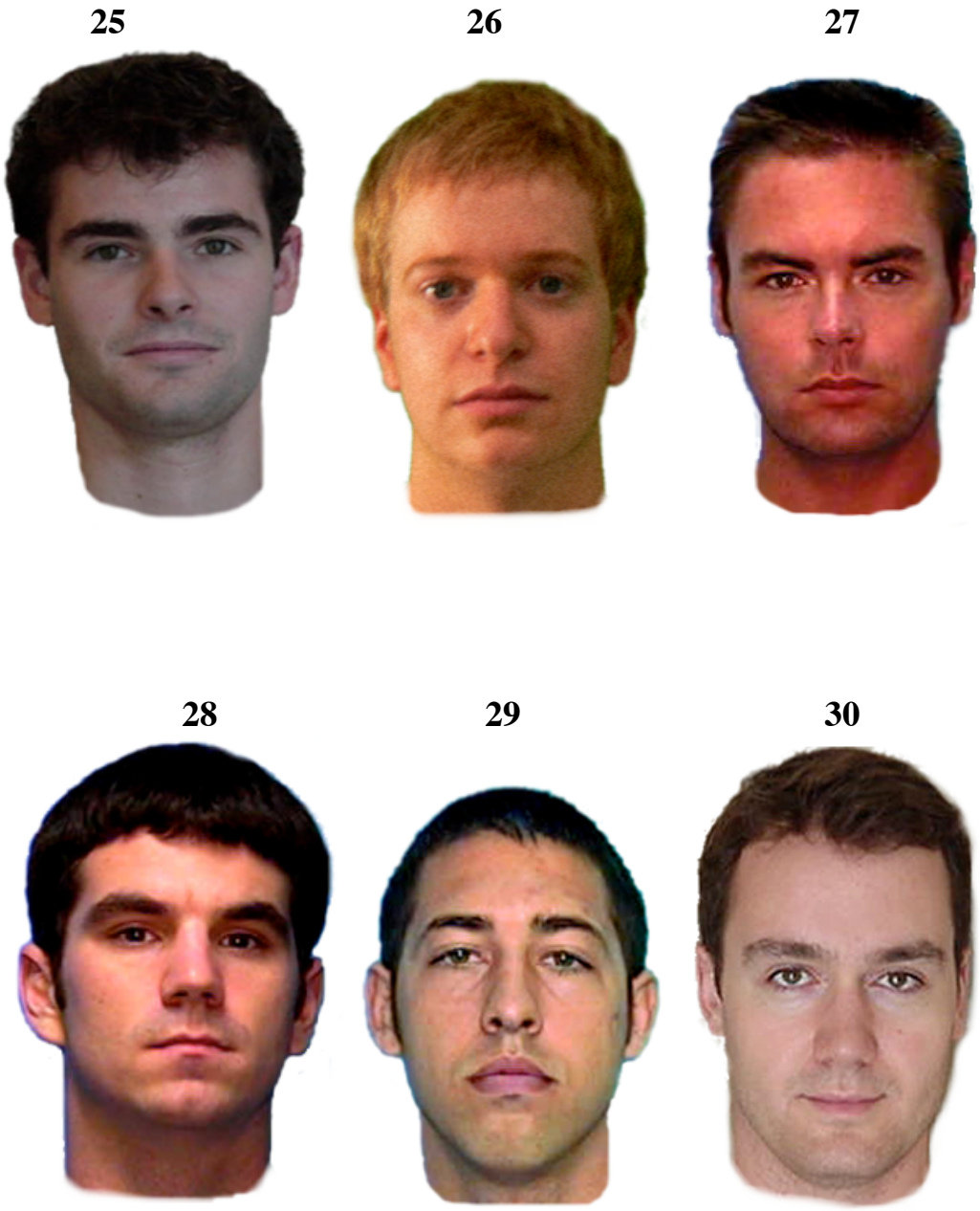
15

16

17

18


19
20
21
22
23
24


25
26
27
28
29
30

31

32

Stimuli Answer Key: Non-Criminal – 1, 2, 6, 7, 9, 12, 13, 14, 15, 17, 18, 19, 22, 25, 26, 30; Arson – 5, 10, 16, 20; Assault – 4, 24, 27, 28; Drug Dealing – 8, 11, 21, 29; Rape – 3, 23, 31, 32