J Nonverbal Behav (2016) 40:255–281
DOI 10.1007/s10919-016-0237-9

ORIGINAL PAPER

Identifying Mental Disorder from the Faces of Women with Borderline Personality Disorder

Alexander R. Daros[1] · Anthony C. Ruocco[1] · Nicholas O. Rule[1]

Published online: 29 June 2016

Abstract Although appearance-based cues can help to diagnose physical illness, visual manifestations of mental disorder may be more elusive. Here, we investigated whether individuals could distinguish women with a serious mental disorder (borderline personality disorder) from demographically- and IQ-matched non-psychiatric controls. Participants rated mentally ill targets as more likely to have a mental disorder from photos more accurately than chance, despite not believing that such judgments were possible. The configuration of facial cues played an important role in these judgments, as interfering with the spatial relationships between facial features reduced participants' accuracy to chance guessing. Further investigation showed similar results when participants rated the targets for specific mental disorders (borderline personality disorder, major depressive disorder) and rated the mentally ill targets as more depressed, angry, anxious, disgusted, emotionally unstable, distressed, and less happy. Moreover, the depression ratings significantly correlated with the targets' actual depressive symptoms. Thus, individuals may be able to infer aspects of mental disorder from minimal facial cues.

Keywords Perception · Facial appearance · Psychopathology · Health · Facial morphology

Introduction

Clinicians regularly rely on patients' subtle and overt nonverbal behavior and appearance to assess mental disorders, noting their facial expressions, vocal tone, eye gaze, physical movements, and emotional range when conducting clinical evaluations (Ekman and Friesen 1974; Garb 2005; Slepian et al. 2014). Despite a long history of speculation about

✉ Alexander R. Daros
alex.daros@mail.utoronto.ca

[1] Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada

the relationship between appearance and mental disorder (Gilman 1982), whether people can accurately identify clinically relevant traits from physical appearance remains a largely untested question. To explore this possibility, we tested whether individuals could reliably discern the mental health of a sample of women diagnosed with borderline personality disorder (BPD) from non-psychiatric controls and investigated several personality traits, clinical characteristics, and physical correlates that might support these perceptions.

Early scholars and philosophers often alluded that disorders of the mind affected one's physical mannerisms and appearance (Davison 2006). Physicians in the late 19th century hypothesized that the physical appearance of individuals with mental disorder contained characteristics that could distinguish them from mentally healthy individuals (Gilman 1982). Today, the idea that minor physical anomalies are more prevalent in individuals with mental disorder is not unfounded. Research has indicated that subtle morphological differences in the craniofacial region can distinguish patients diagnosed with schizophrenia from healthy individuals (Buckley et al. 2005; Lane et al. 1997). Similar studies have shown that patients with other forms of psychosis (McGrath et al. 2002), bipolar disorder (Akabaliev et al. 2011), and autism spectrum disorder (Manouilenko et al. 2014) show higher frequencies of minor physical abnormalities. These features may represent markers of altered morphogenesis during pregnancy related to genetic abnormalities and prenatal events (Waddington et al. 1999), and therefore signifying an increased risk for psychiatric and neurodevelopmental disorders.

A separate and burgeoning literature in social psychology has found that people can extract a surprising amount of information from others' facial appearance (Zebrowitz 1997). Perceivers can discern characteristics such as an individual's sexual orientation (Rule and Ambady 2008), political affiliation (Rule and Ambady 2010), religious beliefs (Rule et al. 2010), and personality traits (e.g., Borkenau and Liebler 1993; Hall et al. 2008; Penton-Voak et al. 2006; Watson 1989) from photographs significantly more accurately than chance. Other work has found that people can reliably infer others' physical health, mental stress, and immunological competence from their faces (Jones et al. 2012; Kramer and Ward 2010; Little et al. 2011). Whether such effects might extend to clinically meaningful judgments remains unknown, however.

Some research has shown that perceivers can reliably distinguish depressed and non-depressed patients (Waxer 1976), individuals with high and low trait anxiety (Waxer 1977), suicidality (Archinard et al. 2000), and various maladaptive personality traits (Fowler et al. 2009; Holtzman 2011; Oltmanns et al. 2004; Vazire and Mehl 2008) based on dynamic behavioral information in videos. More recent studies have suggested that faces alone express cues to depression (Scott et al. 2013), and that people can determine who commits suicide from photos of suicidal and non-suicidal targets' faces (Kleiman and Rule 2013). Scott et al. (2013) created composite faces (digital averages of multiple independent faces) of individuals who had scored high and low on a self-report measure of depression, finding that perceivers could perceive differences in depression from these computer-synthesized faces. Using more natural stimuli (yearbook and family photos), Kleiman and Rule (2013) found that individuals could distinguish people who had died by suicide from matched controls significantly better than chance and determined that perceptions of the targets' impulsive aggression supported this difference (though they did not have information about the individuals' actual impulsivity or aggression to corroborate these perceptions).

Despite the small amount of research and limitations across the extant studies, evidence suggests that people may be able to judge mental health from the face. Here, we tested this hypothesis by systematically determining whether individuals could detect clinically

diagnosed mental disorders from facial appearance. Previous research suggests that dynamic information about an individual can be extrapolated from static photos (Naumann et al. 2009). Whereas research employing videos or brief in vivo interactions may be more accurate, given the increased behavioral information available (Borkenau and Liebler 1992; Waxer 1976), other research suggests that "thin-slices" of behavior from photographs may yield similar levels of accuracy (Kleiman and Rule 2013; Penton-Voak et al. 2006; Scott et al. 2013). In addition, individuals' emotional experiences seem to impact their facial appearance through physiological processes over the lifespan (Ekman 1978; Malatesta et al. 1987). Thus, in the absence of morphological measurements on facial structure, we hypothesized that dynamic information about an individual's emotional and psychological wellbeing might also be extrapolated from facial appearance in static photographs.

Digital photographs were gathered of women diagnosed with (and without) borderline personality disorder (BPD) based on standardized clinical interviews. BPD is a major psychiatric disorder characterized by emotional instability, unstable interpersonal relationships, impulsive aggression, and suicidal behavior (Gunderson 2010; Zanarini and Frankenburg 2007). Other mental disorders, such as mood, anxiety, eating, and substance-use disorders are commonly comorbid in individuals with BPD (Zimmerman and Mattia 1999). Moreover, evidence suggests that symptoms of BPD function as general indicators of distress associated with internalizing (e.g., mood and anxiety), externalizing (e.g., substance use and behavioral dysregulation), and general personality dimensions of psychopathology (Eaton et al. 2011; Sharp et al. 2015). We therefore explored whether untrained observers could identify mental disorder from the faces of women with BPD and investigated characteristics that may support these judgments.

Study 1

To examine the legibility of mental illness from static nonverbal cues in the face, we began by asking participants to rate the faces of women with and without BPD for the likelihood that they suffered from a mental illness in Study 1A. We then honed this investigation by asking participants to rate the faces specifically for the targets' primary (BPD) and most frequent comorbid afflicting illness (major depressive disorder; MDD) in Study 1B, comparing this to judgments of an illness with a direct physiological underpinning (multiple sclerosis).

Study 1A

We examined whether individuals could detect clinically diagnosed mental disorder from faces in Study 1A by photographing women with and without BPD in the laboratory as part of a separate procedure and recruiting separate participants to judge the likelihood that each woman had a mental disorder from her photo.

Method

Stimuli Sixty neutrally-posed individuals who either had a primary diagnosis of BPD ($n = 30$) or no mental disorder at the time of testing ($n = 30$) consented to have their photograph taken during a laboratory visit. We recruited individuals with BPD from an

outpatient clinic at the Center for Addiction and Mental Health and through online and print advertisements in Toronto, Canada, that solicited the participation of individuals with a current diagnosis of BPD. We also recruited healthy individuals from the community through online and print advertisements that solicited adults without a history of mental disorder. We screened all of the candidate targets by phone before completing an in-person assessment. To control for stereotypes about gender and mental illness (e.g., Jones and Cochrane 1981), we only included women. All participants provided a negative urine toxicology screen on the day of testing meaning that they did not have any recreational substances (i.e., cocaine, marijuana, opioids, amphetamine, methamphetamine) in their system. We did not exclude participants for taking their prescribed psychotropic medications (e.g., antidepressants, anxiolytics, mood stabilizers). In addition to providing their photos, all participants also received financial compensation for participating in diagnostic interviewing and completing symptom rating scales.

Our criteria for inclusion required the women to either be mentally healthy or to have a current primary diagnosis of BPD, assessed using the Structured Interview for DSM-IV Personality (SIDP; Pfohl et al. 1997)—a reliable, semi-structured interview designed to assess personality disorders with good interrater reliability and validity (Carcone et al. 2015; Jane et al. 2006). In accordance with the test manual, women were diagnosed with BPD when they met the diagnostic criteria for at least the past 5 years. We assessed comorbid Axis I mental disorders using the Structural Clinical Interview for DSM-IV Axis I Disorders, Patient Edition (SCID-I/P; First et al. 2002). All individuals (including controls) recruited for this study underwent both of the psychodiagnostic interviews. Exclusionary diagnoses included any psychotic disorder, bipolar disorder, neurodevelopmental disorder (e.g., Down's syndrome, autism spectrum disorder), neurological illness, current alcohol or substance use disorder, moderate-to-severe traumatic brain injury, and a significant manual, auditory, or visual impairment. Healthy control targets had no personal history of a DSM-IV mental disorder (either Axis I or Axis II).

Master- and doctoral-level student interviewers trained to reliably administer these measures conducted the diagnostic interviews under the supervision of a licensed clinical psychologist (ACR). Interviewers prepared diagnostic reports for each participant based on all information obtained during the interviews and available medical records, and discussed each case in a best-estimate diagnostic meeting with a licensed clinical psychologist overseeing agreement on diagnosis at which each participant was assigned a Modified Global Assessment of Functioning score (GAF; R. Hall 1995) used by clinicians to rate an individual's overall symptom severity and functional impairment in daily living (see Klein et al. 1994).

The two target groups did not differ in age, ethnicity, intelligence, or years of education (see Table 1). In addition to BPD, the targets had an average of 3.57 ($SD = 2.11$) comorbid Axis I and Axis II mental disorders, with 87 % qualifying for a diagnosis of MDD (see Table 2). None of the participants wore glasses in their photo and roughly equal numbers had visible nose ($n_{\text{BPD}} = 3$; $n_{\text{Healthy Control}} = 5$) and ear piercings ($n_{\text{BPD}} = 11$; $n_{\text{Healthy Control}} = 14$) across the two groups, $\chi^2(1)\text{s} < .14$, $p\text{s} > .60$, $\Phi\text{s} < .04$.

We photographed the targets in the laboratory against the same plain backdrop with consistent lighting while they posed neutral emotional expressions. We then cropped the photos to show just the head and face, standardized them for image height, and auto-adjusted them for brightness and contrast using Adobe Photoshop CS5, v.12.0.5 (see Fig. 1A for sample stimulus). Because attractiveness tends to exert strong effects on judgments from faces (Dion et al. 1972), we asked 39 participants to rate the photos from 1 (*Not at all attractive*) to 7 (*Very attractive*) and averaged their scores for each target

Table 1 Demographic and clinical characteristics of the targets in the present research with statistical tests comparing the two groups

Variable	BPD ($n = 30$)		Healthy controls ($n = 30$)		Test statistic	p
Ethnicity	*N*		*N*			
Caucasian	21		21		$\chi^2(4) = 1.34$	.85
South Asian	4		3			
East Asian	3		2			
Black	1		1			
Latin/mixed	1		3			
	M	*SD*	*M*	*SD*	Test statistic	*p*
Age	32.57	(11.15)	32.73	(11.59)	$t(58) = .06$	.96
Years of education	14.53	(2.43)	14.93	(1.95)	$t(58) = .70$	.49
FSIQ	110.76	(5.93)	110.88	(4.21)	$t(53) = .09$	.93
BDI total	24.86	(9.71)	1.87	(2.68)	$t(57) = 12.49$	<.001
GAF score	46.59	(5.85)	84.37	(2.22)	$t(57) = -33.03$	<.001
BIS composite	4.89	(1.29)	2.05	(.66)	$t(54) = 10.47$	<.001
DERS total	119.96	(27.88)	56.40	(15.02)	$t(53) = 10.77$	<.001

BPD borderline personality disorder, *N* number of individuals, *M* mean, *SD* standard deviation, *FSIQ* full-scale intelligence quotient estimated from the WTAR (Wechsler 2002), *BDI* beck depression inventory-II (Beck et al. 1996), *GAF* global assessment of functioning (Hall 1995), *BIS* Barratt impulsivity scale-11 (Patton et al. 1995), *DERS* difficulty in emotion regulation scale (Gratz and Roemer 2004)

Table 2 Clinical characteristics for individuals diagnosed with BPD in the present research

	M (SD)	N	%
Number of BPD criteria present	6.87 (1.22)		
Number of additional Axis I diagnoses	2.69 (1.49)		
Number of additional Axis II diagnoses	.93 (.94)		
Major depression, current		16	53.33
Major depression, past		26	86.67
Dysthymia, current		1	3.33
PTSD, current		4	13.33
PTSD, past		10	33.33
Alcohol/substance dependence, past		10	33.33
Alcohol/substance abuse, past		4	13.33
Other anxiety disorder, past		13	43.33
Eating disorder, past		7	23.33
Paranoid personality disorder, current		6	20.00
Dependent personality disorder, current		7	23.33
Avoidant personality disorder, current		7	23.33
Narcissistic personality disorder, current		4	13.33
O–C personality disorder, current		3	10.00
Histrionic personality disorder, current		1	3.33

One individual with BPD did not complete all diagnostic assessments. Values were calculated with missing data included

M mean, *SD* standard deviation, *N* number of individuals, % percentage of sample with disorder indicated, *BPD* borderline personality disorder, *PTSD* post-traumatic stress disorder, *O–C* obsessive–compulsive

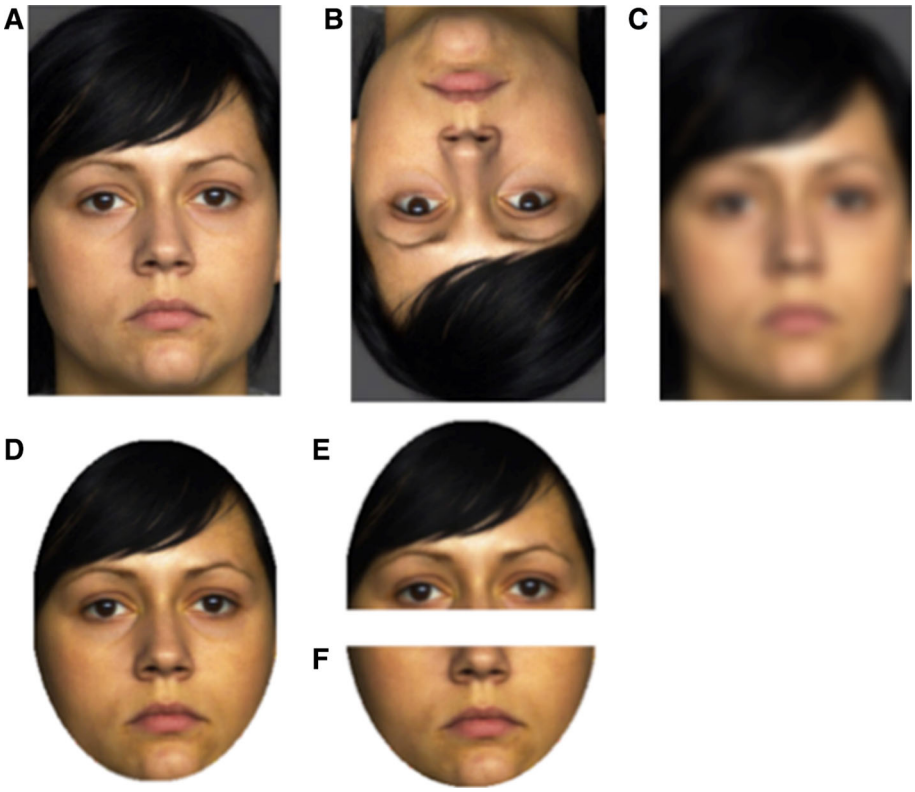

Fig. 1 Sample stimuli used over the course of the studies depicted using an image from a publicly-available stimulus set (FACES; Ebner et al. 2010). Photographs were presented in **a** cropped raw format; **b** inverted; **c** filtered with the Gaussian Blur feature in Photoshop to minimize high spatial frequency visual information; **d** elliptical crop displaying the most prominent internal facial features; **e** upper face; **f** lower face. *(Permission obtained to reproduce image)*

(interrater reliability Cronbach's $\alpha = .98$). Importantly, the targets with mental disorder ($M = 3.30$, $SD = 1.07$) did not significantly differ in attractiveness from the healthy controls ($M = 3.13$, $SD = .88$), $t(76) = .76$, $p = .44$, $d = .17$.

Additional Materials The targets completed several additional measures including the Beck Depression Inventory (BDI-II; Beck et al. 1996) to measure current depressive symptoms, the Difficulties in Emotion Regulation Scale (DERS; Gratz and Roemer 2004) to assess individuals' typical levels of emotion dysregulation, the Barratt impulsiveness scale (BIS-11; Patton et al. 1995) to measure impulsivity, and the Wechsler test of adult reading (WTAR; Wechsler 2002) to estimate premorbid intellectual functioning (IQ) based on each individual's ability to read words of irregular pronunciation (see Table 1 for scores and test statistics for each group). Scores on these measures were used to determine whether perceivers could accurately judge the actual self-reported symptoms reported by each target.

Procedure US residents ($N = 45$; 51 % women; $M_{Age} = 40.61$ years, $SD = 11.9$)[1] were recruited through Amazon's Mechanical Turk (MTurk)[2] to follow a link to an online survey constructed using the Qualtrics survey platform where they viewed the 60 target photographs in random order, rating each for how likely it was that she had a mental disorder from 1 (*Not at all likely*) to 8 (*Extremely likely*).[3] The participants also completed basic demographic information and indicated whether they or anyone they knew had been affected by mental illness. After they rated the targets on the likelihood of having a mental disorder, the participants responded as to whether they believed they could detect mental disorder from facial appearance alone by responding *yes*, *no*, *maybe*, or *unsure* to the question "Do you believe that a mental disorder is something that you can reliably tell about somebody by looking only at their face?" They then selected potentially revealing facial characteristics in response to the question "Which facial features or characteristics might tell you that [a person has] a mental disorder?" from a list of *cheek, chin, dimples, ears, eyebrows, eyelashes, eyes, facial expression, facial symmetry, hairline, jaw, lips, mouth, none, nose, nostrils, skin color, skin complexion, skin quality, wrinkles,* and *something not listed here*. Finally, they answered the open-ended question, "Are there any features of a person's face that might tell you whether they have a mental disorder?"

Results and Discussion

To test the participants' ability to distinguish between women with and without a mental disorder, we analyzed the data using sensitivity correlations by correlating each participant's ratings of the 60 targets with a dummy-coded vector distinguishing the two target groups (Healthy Control = 0, BPD = 1). This produced a point-biserial correlation for each participant, which we converted to a Fisher's z score for inferential analysis. The Fisher's z-scores were then subjected to a one-sample t test that compared the participants' sensitivity correlations to chance (represented by $r = 0$). Overall, participants showed a statistically significant capacity to detect differences in mental illness between the two groups ($M_{\text{Fisher's } z} = .07$, $SD = .11$), $t(44) = 4.27$, $p < .001$, $d = .64$.[4]

[1] All surveys reported in this research took approximately 15–30 min to complete and participants received between $.41 and .91 for their participation (participants serving as targets were paid $100 for their participation as part of a larger neurocognitive study of BPD). The survey contained three blocks: self-reporting demographic information, rating the photographs, and answering additional questions and providing comments. We randomly presented three attention-check questions during the photograph-rating block that consisted of inverted photographs of women borrowed from a separate face database (Ebner et al. 2010). On these trials, participants were asked to select a unique response (e.g., "score of 5") when presented with each stimulus. Participants were included in data analyses only if they had passed all three attention checks (95.1 % of participants across all studies reported in the present work).

[2] MTurk facilitates high quality data collection from a large pool of diverse participants by allowing job requesters to reject participants' work if they do not follow instructions. Validation studies have found that participants are highly motivated to complete the tasks even when low hourly compensation rates (Buhrmester et al. 2011; Peer et al. 2014) and that individuals who complete online measures on MTurk tend to be very similar to those who physically come to the lab (Goodman et al. 2013).

[3] See "Appendix 1" for tests of variation according to participant and target demographics in all of the studies reported here and "Appendix 2" for detailed participant demographics for each study.

[4] Notably, dichotomizing the 8-point scale to analyze the data using signal detection analyses (Macmillan and Creelman 2005) produced similar results: participants categorized the targets' mental health status significantly more accurately than chance ($M_{A'} = .56$, $SD = .11$), $t(44) = 3.43$, $p = .001$, $d = .51$, and showed a non-significant tendency towards assuming that targets were healthy, rather than mentally ill ($M_{B''} = .07$, $SD = .37$), $t(44) = 1.36$, $p = .18$, $d = .20$, consistent with the prevalence of mental illness in society (Kessler et al., 2005).

Roughly half (51 %) of the participants thought that it was not possible to perceive mental disorder from facial appearance alone, 42 % chose *maybe or unsure*, and 7 % chose *yes*. Effect-coding the participants' responses (1 = Yes, 0 = Unsure, −1 = No) showed that perceivers' beliefs about the legibility of mental illness did not significantly relate to their accuracy ($r = .20$, $p = .18$)[5]. In contrast, only 11.1 % of the same sample indicated that no facial characteristics could serve as indicators of mental disorder (i.e., selected *none*). The top characteristics selected by participants were: *facial expression* (51.1 %), *eyes* (30.0 %), *facial symmetry* (28.9 %), *mouth* (24.4 %), *eyebrows* (22.2 %), *skin complexion* (17.8 %), *skin quality* (17.8 %), and *lips* (15.6 %). Few participants selected *something not listed* as an indicator (8.9 %). Thus, the participants showed little consensus about what facial characteristics they believed could indicate a mental disorder, and their beliefs about the detectability of mental disorder from faces did not predict their sensitivity to it.

Responses to the open-ended questions paralleled those of the closed-ended questions, suggesting that the participants were generally unaware that they could detect mental disorder from individuals' faces. Answers ranged from socially-desirable abstentions (e.g., "[it is wrong] to judge a book by its cover") and responses consistent with the majority closed-ended response above (e.g., "it is impossible to tell for sure"), to uncertainty stemming from nuance (e.g., "the appearance of someone cannot [lead] you to conclude that they have a mental illness… But sometimes [extreme differences] can be implied") and acceptance of the possibility, often supported by specific lay theories about mechanisms ("overall, people that I meet who have a mental disorder have a sad or vacant look in their eyes").

Study 1B

In Study 1A, participants evaluated the targets using the general label "mental disorder." Given that the targets all shared a common primary diagnosis of BPD, and as most also suffered from comorbid MDD, we narrowed our investigation in Study 1B by asking participants to rate these disorders, specifically. In addition, we asked separate participants to rate targets on the likelihood each had a neurological condition called multiple sclerosis to investigate whether the targets with a mental disorder were perceived to have a physical illness more often than those with no psychiatric history. Here, we expected non-significant differences between these likelihood ratings, suggesting that the degree of physical illness discerned from faces did not influence ratings of mental disorder.

Method

A total of 120 American (i.e., US resident) MTurk Workers either rated the likelihood that each of the 60 targets had BPD ($N = 40$; 57.5 % women; $M_{Age} = 34.4$ years, $SD = 13.1$), MDD ($N = 42$; 38.1 % women; $M_{Age} = 35.3$ years, $SD = 12.6$), or multiple sclerosis ($N = 38$; 68.4 % women; $M_{Age} = 36.1$ years, $SD = 13.6$) using analogous scales and procedures as in Study 1A.

[5] Additional one-sample *t*-tests according to participants' belief category showed that those who did not believe mental illness to be legible from faces differentiated the targets significantly better than chance ($M_{Fisher's\ z} = .10$, $SD = .09$), $t(22) = 5.13$, $p < .001$, $d = 1.05$, whereas those who were unsure ($M_{Fisher's\ z} = .04$, $SD = .13$), $t(18) = 1.34$, $p = .20$, $d = .30$, and those who responded "yes" ($M_{Fisher's\ z} = .07$, $SD = .09$), $t(2) = 1.30$, $p = .32$, $d = .53$, did not; though the very small number of participants who believed that mental illness was legible from the prevented meaningful interpretation.

Participants randomly assigned to the BPD condition first read a paragraph listing the diagnostic criteria for BPD, explaining that the presence of five of the nine diagnostic criteria satisfies a diagnosis of BPD. Similarly, participants randomly assigned to the MDD condition read over the diagnostic criteria for MDD and rated each target for whether she seemed "currently depressed". Finally, the remaining participants read a description of multiple sclerosis as "a medical disorder in which motor function is affected by damaged neuronal insulation" along with several symptoms associated with the disorder (e.g., muscle weakness, fatigue, problems in speech and vision, cognitive slowing) and rated each woman for the likelihood that she suffered from the disease.

Results

A series of one-sample *t*-tests demonstrated that participants detected the targets with mental illness significantly better than chance when rating them for BPD ($M_{\text{Fisher's } Z} = .11$, $SD = .11$), $t(39) = 6.22$, $p < .001$, $d = 1.00$, and MDD ($M_{\text{Fisher's } Z} = .07$, $SD = .09$), $t(41) = 4.66$, $p < .001$, $d = .73$. In addition, participants' ratings of BPD significantly correlated with the targets' self-reported emotional instability on the DERS ($M_{\text{Fisher's } Z} = .10$, $SD = .12$), $t(39) = 5.09$, $p < .001$, $d = .82$, and participants' ratings of MDD significantly correlated with the targets' self-reported depression on the BDI ($M_{\text{Fisher's } Z} = .07$, $SD = .09$), $t(41) = 5.27$, $p < .001$, $d = .82$. The effect sizes obtained from participants' ratings of "mental disorder" in Study 1A did not significantly differ in magnitude from those obtained from participants' ratings of BPD and MDD, ts < 1.67, ps $> .10$, ds $< .36$, when compared both with independent *t*–test statistics and meta-analytically (presented in Table 3 along with a list of results by study).

The mean likelihood ratings for multiple sclerosis did not differ significantly between targets with a mental disorder ($M = 3.98$, $SD = .76$) and healthy controls ($M = 3.96$, $SD = .73$), $t(58) = .12$, $p = .91$, $d = .03$. Corroborating this finding, there was a non-significant finding for the sensitivity analysis associated with distinguishing the target sample using a description of multiple sclerosis ($M_{\text{Fisher's } Z} = -.00$, $SD = .12$), $t(37) = .10$, $p = .92$, $d = -.02$. While the between-group differences for mean likelihood of "mental disorder", BPD, and MDD were also non-significant (ts < 1.99, ps $> .052$, ds $< .51$), the effect size magnitude was comparatively lower for multiple sclerosis. Collectively, these analyses suggest that targets with a mental disorder were not perceived to have a physical illness to a greater degree than healthy individuals overall.

Discussion

Similar to studies examining the legibility of personality traits and social characteristics from photos of faces, participants here distinguished women with BPD from non-psychiatric controls. These findings echo those of previous research examining the identification of depression and suicidality from facial cues. Scott et al. (2013) reported some evidence for the visibility of depression in the face, and Kleiman and Rule (2013) found that perceivers could distinguish those who committed suicide from matched controls. The present study adds to this literature by demonstrating that perceivers can detect mental disorder from the faces of women with a primary diagnosis of BPD significantly better than chance guessing.

Participants showed similar sensitivity to detecting mental disorder when rating the specific illnesses most strongly characterizing the targets (i.e., BPD and MDD) but not when rating an unrelated physical illness as a control (multiple sclerosis). Thus, judgments

Table 3 Descriptive statistics in distinguishing between targets with borderline personality disorder and healthy controls according to study and condition

Study	Rating condition	*N*	$M_{\text{Fisher's z}}$	*SD*	95 % *CI*	*Z*-value
1A	Mental disorder	45	.07	.11	[.04, .10]	4.27
1B	Borderline personality disorder	40	.11	.11	[.08, .14]	6.33
	Major depressive disorder	42	.07	.09	[.04, .10]	5.04
	Multiple sclerosis	38	.00	.12	–	–
2A	Depressed	31	.09	.10	–	–
	Anxious	31	.07	.12	–	–
	Impulsive	32	.04	.11	–	–
	Emotional instability	35	.07	.12	–	–
	Introversion	30	−.04	.11	–	–
2B	Anger	43	.11	.17	–	–
	Fear	43	.04	.19	–	–
	Disgust	43	.13	.16	–	–
	Happiness	43	−.19	.15	–	–
	Sadness	43	.02	.16	–	–
	Surprise	43	.01	.15	–	–
	Distressed	37	.13	.13	–	–
3A	Inverted	30	.00	.12	[−.04, .04]	.00
	Blurred	33	.05	.08	[.02, .08]	3.59
3B	Elliptically-cropped faces	43	.07	.14	[.03, .11]	3.28
3C	Upper face	52	.03	.11	[.00, .06]	1.97
	Lower face	52	.01	.11	[−.02, .04]	.66
Meta-analytic summary of accuracy data[a]		337	.05[b]	.04[c]	[.04, .06]	9.03

Z-value represents the number of standard deviations the obtained mean is above or below the meta-analytic mean related to its standard error

M mean, *SD* standard deviation, *N* number of individuals, *CI* confidence interval

[a] A meta-analytic summary of accuracy in detecting "mental disorder" (as well as specific mental disorders) from facial appearance was conducted using the sensitivity data from Studies 1 and 3 (excluding the multiple sclerosis control condition)

[b] $Q(7) = 30.0$, $p < .001$

[c] The standard error from the meta-analytic output was .01

of both general and specific mental disorder distinguished the two groups. Using ratings of only one physical illness limits generalization of this null effect to other physical ailments, however. Additional analyses showed that participants accurately inferred targets' levels of depression from their faces, extending previous work by Scott et al. (2013) that documented perceivers' ability to discriminate between computer-synthesized composites of individuals who scored either high or low on a self-report measure of depression to the faces of real individuals.

These findings accord with previous research demonstrating that participants can accurately judge physical health, mental stress, and immunological competence from facial appearance (e.g., Jones et al. 2012; Kramer and Ward 2010; Little et al. 2011). Yet earlier studies on the accurate perception of mental illness exhibited far less control. Scott et al.

(2013) used computer-synthesized composites of individuals' faces, and Kleiman and Rule (2013) used faces taken from external sources without much information about the targets. In contrast, the present study used a rich and highly controlled set of real stimuli: individuals sat for photographs under standardized conditions in the laboratory and completed a battery of diagnostic tests, producing extensive demographic and symptom-related information about them.

The majority of participants in Study 1A expressed disbelief that mental disorder could be expressed or perceived via facial cues but, in contradiction, also identified characteristics of facial appearance that might possibly contain cues to mental health. Moreover, they judged mental disorder better than chance. This disconnect between what people do and what they believe suggests that the process by which they extract information about mental disorder from facial appearance may occur outside of conscious awareness (see Rule et al. 2008, for parallel findings for judgments of sexual orientation). To account for other characteristics of the targets that might cue differences between them, we explored participants' perceptions of a variety of BPD correlates in Study 2.

Study 2

Participants in Study 1 reliably distinguished between women with and without a mental disorder from photos of their faces. Here, we sought to better understand some of the factors that might have supported these judgments. In Study 2A, we hypothesized that perceivers might infer target individuals' mental health status by perceiving traits associated with BPD (similar to how perceptions of impulsivity informed ratings of suicide completion in previous work; Kleiman and Rule 2013). Additionally, although the targets presented neutral expressions in their photographs, perceivers might nonetheless attempt to infer mental disorder from subtle cues to negative affect (e.g., sadness, fear, and anger; see also Malatesta et al. 1987), given the well-known association between negative affect with mental illness (Clark et al. 1994) Thus, participants rated each target on six basic emotional expressions in Study 2B.

Study 2A

Here, independent samples of perceivers rated the targets on one of several behaviorally- and clinically-relevant characteristics found to typically associate with BPD in previous research: depression, anxiety, emotional instability, social avoidance (tested as "introversion"), and impulsivity (Glenn and Klonsky 2009; Gratz et al. 2013; Gunderson 2010; Koenigsberg et al. 2002; Trull et al. 2008; Zimmerman and Mattia 1999).

Method

A total of 159 American MTurk Workers (53 % women; $M_{Age} = 35.0$ years, $SD = 11.7$) rated the faces for either depression ($n = 31$), anxiety ($n = 31$), emotional instability ($n = 35$), introversion ($n = 30$), or impulsivity ($n = 32$). Participants rated the same 60 target photos as in Study 1 from 1 (*Not at all* [trait]) to 7 (*Very* [trait]) following procedures similar to those described above. As in Study 1A, they then indicated whether they believed they could detect mental illness from facial appearance using the same series of closed- and open-ended questions.

Results

Sensitivity correlations between the individual participants' ratings and the targets' mental health status (Healthy Control = 0, BPD = 1) showed that they rated the mentally ill targets significantly higher than the healthy controls on depression ($M_{\text{Fisher's } z} = .09$, $SD = .10$), $t(30) = 5.01$, $p < .001$, $d = .91$, emotional instability ($M_{\text{Fisher's } z} = .07$, $SD = .12$), $t(34) = 3.42$, $p = .002$, $d = .59$, and anxiety ($M_{\text{Fisher's } z} = .07$, $SD = .12$), $t(30) = 3.52$, $p = .001$, $d = .64$, but not impulsivity ($M_{\text{Fisher's } z} = .04$, $SD = .11$), $t(31) = 2.45$, $p = .02$, $d = .44$, or introversion ($M_{\text{Fisher's } z} = -.04$, $SD = .11$), $t(30) = 1.92$, $p = .06$, $d = -.36$, when accounting for multiple comparisons (Bonferroni-corrected $\alpha = .01$; see Appendix 3 for descriptive statistics). Comparing the sensitivity correlations between the samples of participants whose judgments differentiated between the target groups (i.e., those who rated depression, emotional instability, and anxiety) did not reveal significant differences in the magnitude of the effects, suggesting that these traits distinguished the targets to a similar extent (all pairwise $|t|$s $\leq .73$, ps $\geq .47$).

Given that the targets completed a variety of relevant measures at intake, we additionally measured how well the participants' perceptions matched the targets' self-reported depression, impulsivity, and emotion regulation difficulties. Sensitivity correlations between each participant's perceptions of depression, impulsivity, and emotional instability and the targets' scores on the BDI, BIS, and DERS, respectively showed legibility for depression ($M_{\text{Fisher's } z} = .09$, $SD = .10$), $t(30) = 4.84$, $p < .001$, $d = .88$, but neither impulsivity ($M_{\text{Fisher's } z} = .03$, $SD = .12$), $t(31) = 1.38$, $p = .18$, $d = .25$, nor emotional instability ($M_{\text{Fisher's } z} = .03$, $SD = .15$), $t(34) = 1.28$, $p = .22$, $d = .22$.

Finally, the majority of perceivers (54.7 %) across all conditions did not believe that they could perceive mental disorder from facial appearance (8.2 % responded *yes* and 37.1 % responded *maybe/unsure*). Similar to Study 1A, however, only 20.8 % of the participants maintained this opinion when selecting facial characteristics that could serve as possible indicators of mental disorder (i.e., by selecting *none*). The top choices of possible cues were: *eyes* (71.7 %), *facial expression* (57.9 %), *mouth* (32.1 %), *something not listed here* (20.8 %), *facial symmetry* (18.9 %), *lips* (17.0 %), *eyebrows* (15.1 %), and *skin complexion* (13.8 %).

Study 2B

Given that emotions affect the physical configuration of facial appearance through the autonomic nervous system (Ekman 1978), facial appearance may come to communicate information to observers about an individual's chronic affective experience over time (Hess et al. 2012; Malatesta et al. 1987). That is, people with particular dispositions (e.g., anxiety) may come to develop static facial appearances that subtly reflect their frequent but fleeting emotional states (e.g., an anxious person may come to look permanently fearful). Although our targets posed neutral expressions when photographed, we therefore investigated whether individuals might nevertheless perceive subtle signals of emotions from their faces and whether these might systematically differ between the patients with mental illness and the healthy controls.

Method

American MTurk Workers ($N = 43$; 51.2 % women; $M_{\text{Age}} = 35.9$ years, $SD = 13.0$) rated each of the 60 targets on six basic expressions of emotion (anger, fear, disgust,

happiness, sadness, and surprise) from 1 (*Not at all* [emotion]) to 7 (*Very* [emotion]) in a within-subjects design. Procedures followed those described above except that participants rated all of the emotions simultaneously for each face. To capture further complexity in the targets' potential emotional states, separate MTurk workers ($N = 37$; 43.2 % women; $M_{Age} = 34.5$ years, $SD = 12.5$) rated the degree of the targets' apparent distress from 1 (*Not at all distressed*) to 7 (*Very distressed*).

Results

On average, participants perceived the mentally ill targets as significantly angrier ($M_{Fisher's\ z} = .11$, $SD = .17$), $t(42) = 4.11$, $p < .001$, $d = .64$, more disgusted ($M_{Fisher's\ z} = .13$, $SD = .16$), $t(42) = 5.44$, $p < .001$, $d = .85$, and less happy ($M_{Fisher's\ z} = -.19$, $SD = .15$), $t(42) = 7.92$, $p < .001$, $d = -1.22$, than the healthy controls but showed no distinction between the groups in their ratings of sadness ($M_{Fisher's\ z} = .02$, $SD = .16$), $t(42) = .97$, $p = .34$, $d = .15$, surprise ($M_{Fisher's\ z} = .01$, $SD = .15$), $t(42) = .50$, $p = .62$, $d = .08$, or fear ($M_{Fisher's\ z} = .04$, $SD = .19$), $t(42) = 1.37$, $p = .18$, $d = .21$. Participants evaluating the targets' apparent distress rated the BPD patients as significantly more distressed than the healthy controls ($M_{Fisher's\ z} = .13$, $SD = .13$), $t(36) = 5.74$, $p < .001$, $d = .96$. Additional sensitivity correlations between their distress ratings and the targets' GAF scores ($M_{Fisher's\ z} = -.13$, $SD = .13$), $t(36) = 6.63$, $p < .001$, $d = -1.03$, showed that perceptions of greater distress corresponded to more compromised functioning, demonstrating convergence between the impressions of our naïve participants and those of trained clinicians.

Discussion

Participants in Study 1 showed sensitivity to differences in mental illness communicated by women's faces. We built upon that finding in Study 2A by identifying traits relevant to BPD and mental disorder that distinguished the target groups. Although participants perceived women with mental disorder to look more anxious, depressed (as in Study 1B), emotionally unstable, impulsive, and less introverted than women without a mental disorder, the magnitude of these results reached statistical significance only for the first three traits. The effects sizes for these three traits did not significantly differ, suggesting that none of anxiety, depression, or emotional instability dominates distinctions between the two groups. However, additional analyses showed that participants accurately inferred targets' levels of depression from their faces, as in Study 1B.

Perceptions of negative emotional expressions from targets' neutral faces may also reveal their mental health status. Participants in Study 2B rated targets with BPD as significantly more unhappy, disgusted, and angry than the healthy controls. Although one might expect apparent sadness to distinguish the two groups (consistent with the differences in perceived and actual depression reported above), the participants did not perceive the mentally ill targets as any sadder than the controls. A separate group of participants did perceive the BPD patients as looking significantly more distressed, however.

Finally, similar to the results of Study 1A, participants again displayed a discrepancy between what they did and what they said regarding the ability to detect mental disorder from facial appearance, suggesting that they may lack awareness about their ability to reliably perceive cues to mental illness from the face. Overall, the results of Study 2 therefore provided some insight as to the behavioral or trait-based inferences that perceivers might make when judging mental illness. Although these findings may speak to

why participants perceive some faces to have a mental disorder, they do not answer the question of *how* participants perceive the faces to make their judgments. We thus examined the facial characteristics that perceivers use to evaluate mental illness in Study 3.

Study 3

To better understand what characteristics of targets' faces perceivers use to accurately infer mental disorder, we explored judgments based on specific facial cues in Study 3. At its broadest level, face perception typically occurs through processing either individual facial characteristics (e.g., eyes, nose, mouth; referred to as *featural processing*) or through processing the face as a unified whole or configuration (*configural processing*; see Maurer et al. 2002, for review). Information conveyed by individual facial characteristics should thus occur through featural processing whereas information encoded in the interrelationships between separate characteristics would rely on configural processing.

Classic research has shown greater complexity for processing faces presented upside-down (Yin 1969). For judgments that rely on the configuration of facial characteristics, inverting the faces impairs perceptual processing because it disrupts the relationships between the face's physical characteristics. The holistic perception of faces therefore becomes obstructed and the perceiver must reconstruct the face through featural processing: extracting each feature piece-by-piece and then cognitively reassembling them into a whole face. This slows the processing of faces because it requires extra steps and can also undermine accuracy when the outcome relies on specific aspects of the configuration (Tanaka and Farah 1993). Thus, if accurate perceptions of mental disorder from faces rely on configural processing, inverting the targets' faces should hinder the accuracy of perceivers' judgments.

Alternatively, face inversion does not typically affect featural processing. Because individual facial characteristics provide information independently, the spatial relationships between one feature and the rest of the face do not weigh critically in rendering an accurate judgment. Disrupting the configuration (as through inversion) does not obstruct the extraction of information from individual features, then, and judgmental accuracy should persevere. The properties of specific facial characteristics are important, however. Obscuring the details of the features of an upright face (as through blurring it) therefore challenges featural processing while leaving the configuration of the face relatively intact (Goffaux et al. 2005).

To test which of these two processes might underlie accurate perceptions of mental disorder, we transformed the photos from Studies 1 and 2 by either inverting or blurring them to disrupt perceivers' configural and featural processing of the images, respectively. Accurate judgments of inverted (but not blurred) faces would suggest that featural processing supports the accuracy of participants' judgments of mental disorder. Alternatively, accurate judgments of blurred (but not inverted) faces would suggest that configural information underlies the accuracy of perceivers' judgments. Thus, we investigated the role of configural and featural processing in perceivers' accurate judgment of mental disorder in Study 3A.

To follow up on this, we examined perceivers' judgments of elliptically-cropped versions of the original images from Studies 1 and 2 (thereby presenting only the targets' internal facial characteristics) in Study 3B. Should participants continue to show accuracy without viewing extraneous features such as hairstyles and clothing, it would confirm the

importance of facial characteristics (vs. external cues) in perceivers' inferences of mental disorder. Finally, to further clarify the respective roles of configural processing in perceivers' accurate judgments of targets' mental disorder, we split the elliptically-cropped faces from Study 3B to display only the upper or lower portion of the faces. Perceivers in Study 3C then judged either the upper or lower halves of the faces. Should configural processing be vital to the accurate judgment of mental disorder in targets, separating the faces into halves would prevent perceivers from accurately judging targets on the basis of mental disorder. This would then suggest that information from the upper and lower portion of the face, related to the configuration of facial characteristics as a whole, is required to support inferences of mental disorder.

Study 3A

Method

To investigate whether accurate judgments of mental disorder from faces rely primarily on configural or featural processing, we created two experimental conditions in which perceivers completed the same task as in Study 1 but with either inverted or blurred versions of the 60 target faces. In the first condition, we rotated the faces 180° along the vertical axis so that each appeared upside-down (Fig. 1b). In the second condition, we distorted each face using a Gaussian blur to lower the spatial frequency of the photographs in a way that preserved the internal facial configuration but drastically reduced the details of the facial characteristics (Fig. 1c). American MTurk Workers then rated either the inverted ($N = 30$, 63 % women; $M_{\text{Age}} = 34$ years, $SD = 12.8$) or blurred ($N = 33$, 42 % women; $M_{\text{Age}} = 36.5$ years, $SD = 12.1$) faces following the same procedures as in Study 1.

Results

Perceivers' sensitivity correlations showed that participants distinguished targets with and without mental disorder significantly better than chance when judging the blurred faces ($M_{\text{Fisher's } Z} = .05$, $SD = .08$), $t(32) = 3.22$, $p = .003$, $d = .56$, but not the inverted faces ($M_{\text{Fisher's } Z} = .00$, $SD = .12$), $t(29) = .04$, $p = .97$, $|d| < .01$. Participants may therefore rely more on the relationships between facial characteristics to judge mental illness among patients with BPD than on any individual feature itself. Direct comparison of perceivers' accuracy when judging the inverted versus blurred faces showed only a marginally significant difference, however: $t(61) = 1.96$, $p = .054$, $d = .50$. Thus, a face's configuration may be more valuable than its specific characteristics when perceivers are extracting information about mental disorder; however, other aspects may still be involved. We addressed this possibility in Studies 3B and 3C.

Study 3B

Method

To examine perceivers' ability to infer mental disorder independent of extra-facial characteristics, we elliptically cropped the photos from Studies 1 and 2 to remove peripheral features such as photo backgrounds, hairstyles, and face shape (Fig. 1D) while retaining

the face's internal characteristics.[6] American MTurk Workers ($N = 43$, 58 % women; $M_{Age} = 36.1$ years, $SD = 14.4$) then judged mental disorder based on these elliptically cropped images following the same procedures as above.

Results

Removing extra-facial aspects had little effect on the detection of mental disorder. Participants' sensitivity to mental illness based on the elliptically cropped photos continued to significantly exceed chance ($M_{Fisher's\ Z} = .07, SD = .14), t(42) = 3.08, p = .004, d = .47$.[7]

Study 3C

Method

The comparison between perceivers' accuracy for the blurred and inverted faces in Study 3A only reached marginal levels of statistical significance. We therefore asked American MTurk Workers to judge either the upper ($N = 52$; 50 % women; $M_{Age} = 35.2$ years, $SD = 10.6$; Fig. 1E) or lower ($N = 52$; 25 % women; $M_{Age} = 39.8$ years, $SD = 16.6$; Fig. 1f) halves of the elliptically cropped faces in isolation to further assess whether configural or featural cues support the accuracy of perceivers' judgments. Procedures were the same as those described above.

Results

Perceivers judging neither the upper ($M_{Fisher's\ Z} = .03$, $SD = .11$), $t(51) = 1.89$, $p = .07$, $d = .26$, nor lower ($M_{Fisher's\ Z} = .01$, $SD = .11$), $t(51) = .81$, $p = .42$, $d = .11$, portions of the elliptically cropped faces judged mental disorder significantly more accurately than chance guessing. These data therefore underscore the importance of retaining the face's intact configuration to successfully judge mental disorder.

Discussion

Overall, the results of Study 3 suggest that perceivers rely on configural facial cues to judge mental illness. In Study 3A, blurring the details of the faces' physical characteristics did not obstruct participants' accurate judgment of mental illness but inverting the faces did. Consistent with this, presenting participants with just the upper or lower portions of the face did not provide sufficient information for them to make their judgments—emphasizing the need to judge mental illness based on the whole intact face to achieve accuracy. Finally, removing extrafacial information (such as that from hairstyle and face shape) did not result in lower levels of accuracy, underlining the role of the face's internal

[6] Some cropped photographs (37 %) still contained visible hair (e.g., bangs) but this did not vary in frequency between the two target groups ($n = 11$ in both groups).

[7] Perceivers in Study 3B also provided feedback about potential facial features that might reveal whether an individual has a mental disorder, as in Studies 1 and 2. Results converged with those reported above: most perceivers (61 %) reported that they could not detect mental disorder from facial appearance, 37 % responded *maybe/unsure*, and 2 % responded *yes*. Nevertheless, only 14 % of participants affirmed their doubt about mental disorder's legibility when selecting possible revelatory features (14 %). Top choices consisted of *eyes* (79.1 %), *facial expression* (65.1 %), *mouth* (32.6 %), *facial symmetry* (25.6 %), *lips* (16.3 %), *eyebrows* (16.3 %), *skin complexion* (16.3 %), *skin quality* (20.9 %), and *not listed* (4.7 %).

characteristics in perceivers' judgments. Altogether, these findings lend greater understanding as to how perceivers extract information about mental illness from targets' faces.

General Discussion

Contrary to past work examining the relationship between facial appearance and mental health (e.g., Kleiman and Rule 2013; Scott et al. 2013), our study uniquely used targets that underwent structured diagnostic interviewing to comprehensively assess the presence of mental disorder. Half of the target individuals had BPD, a major mental disorder that affects emotion regulation, impulse control, and interpersonal functioning (Gunderson 2010). Based on these targets, the present results suggest that information in facial appearance allows people to distinguish between groups of individuals diagnosed with and without mental disorder.

Participants asked to rate the faces of women with and without BPD showed better than chance sensitivity in detecting differences between the two target groups when making general inferences of "mental disorder". Participants showed similar sensitivity when judging the disorders that most strongly characterized the targets' mental health (i.e., BPD and MDD) but not when judging an unrelated physical illness (i.e., multiple sclerosis). Targets with a mental disorder thus did not simply look physically ill overall, suggesting that there are observable differences between physical and mental health. Significant relationships between perceivers' inferences of depression and BPD and the targets' actual measured levels of depression and emotional instability, respectively, further validated these results. Thus, women's faces seem to contain subtle information about their mental health that perceivers can detect from their appearance.

Participants also perceived the women with mental disorder as more depressed, emotionally unstable, anxious, impulsive, and less introverted than the healthy controls. Although these differences were greater for the first three traits, all of the traits tested showed at least small effect size differences between the groups (see Cohen 1988). However, these differences only emerged when we used sensitivity analyses and none of the group means significantly differed when aggregating the scores for each target. Moreover, only perceptions of depression corresponded to the targets' actual depression levels in further analyses. Neither perceptions of trait impulsivity nor trait emotional instability correlated with targets' scores on the scales assessing these traits (i.e., the BIS and DERS, respectively). These non-significant findings may suggest that the relationship between the clinical tools used to assess the traits did not match well with the questions we asked the perceivers. Nevertheless, these results help to map the landscape of what clinical characteristics may be legible from facial appearance and how they correspond to individuals' actual self-reported behavior.

In addition, participants rated the patient sample as significantly angrier, more disgusted, and lower in apparent happiness than the healthy controls, despite their neutral expressions (though the two groups appeared similarly sad, surprised, and afraid). The participants also rated the BPD patients as significantly more distressed compared to healthy controls, and these ratings correlated with each individual's clinically-rated assessment of global functioning. Collectively, these results suggest that emotional facial expressions may also influence the likelihood of applying the label of mental disorder. Individuals with BPD may display more negative affect and therefore look less happy. They also tend to struggle with intense and inappropriate anger as well as self-criticism (Gunderson 2010), which might relate to the perception of greater anger and disgust among people with BPD in the present work.

In addition to these trait and expressive differences between the two groups, we also examined what spatial-physical characteristics of the face supported inferences of mental illness. Disrupting the face's configuration by inverting it or showing only its upper or lower portions interfered with participants' ability to reliably infer mental disorder. In contrast, blurring the faces' internal features did not impair the legibility of mental disorder. Thus, the relationships among the face's features seem to support accurate inferences of mental disorder, rather than the details of the features themselves.

A previous survey found that 78 % of community-dwelling respondents felt that they could "tell if someone was mentally ill," though this also included cues from behavior, speech, style of dress, facial expression, and personality traits (Wolff et al. 1996). In stark contrast, the vast majority (i.e., more than 85 %) of the perceivers that we sampled reported that they thought themselves unable to reliably detect cues to mental disorder from a person's face. Despite the perceivers' general disbelief that they could discern mental disorder from facial appearance, however, three-quarters of them simultaneously reported the belief that areas around the eyes and mouth (as well as facial expression, symmetry, and complexion) provide important cues to mental disorder. Thus, perceivers' tendency to deny being able to judge mental disorder from facial cues might actually stem from socially-desirable responding, whereby individuals may think it impolite to concede that cues to stigma are present and perceptible (e.g., Apfelbaum et al. 2008).

Previous research has shown that perceptions of impulsivity and emotional stability from faces are not only distinct (Penton-Voak et al. 2006) but may predict important outcomes (Kleiman and Rule 2013). Given the dimensional nature of personality and psychopathology, however, each individual may vary in his or her combination of clinical symptoms and character traits, which could introduce noise that prevents the observation of a clear signal from particular comorbid traits (Watson et al. 2008). Indeed, participants consistently rated some of the mentally ill targets as having a mental disorder more often than others; perceptions of other traits (measured and not measured) may have influenced these results. The present targets may therefore not allow for the best test of detecting these correlated traits (e.g., introversion). Further research should therefore fully vet the possible legibility of these other traits with other samples to better understand the relationship between facial appearance and the various clinical characteristics associated with mental disorder.

Limitations

One should consider several limitations when interpreting the results of these studies. First, although the findings indicate accuracy in judging mental disorder, the results exceed chance by only a small margin. Thus, despite perceivers' ability to accurately perceive mental disorder, mistakes in judgment occurred frequently as well. Indeed, whereas the findings provide insight into what facial characteristics individuals may use to infer the presence of mental disorder, face-based judgments should not supplant conventional clinical diagnostic methods of assessment. Moreover, the results do not address questions about the facial cues that may potentially signify the presence of mental disorder but, rather, a given individual's ability to detect them. Despite the many laudable qualities about the target sample in terms of matching, diagnosis, and control, we lacked a large enough sample to reliably measure the physical dimensions distinguishing the two groups. Relatedly, the stark differences in well-being between our healthy and patient samples made it difficult to determine whether participants were attuned to clinically significant levels of mental disorder, or to general mental distress and maladjustment. Nevertheless, if

judgments of mental disorder are accurate at least some of the time, then understanding how individuals in society use this information may be important for both basic and applied understanding of psychopathology.

This research also focused on a single mental disorder: BPD. Despite this narrow focus, BPD represents broader mental illness well due to its extensive psychiatric diagnostic comorbidity and functional disability (Grant et al. 2008; Ruocco et al. 2014), clearly reflected among the targets photographed for the current work. Research has also indicated a relationship between BPD symptoms and psychopathology in general (Eaton et al. 2011; Sharp et al. 2015). Centering on a primary diagnosis of BPD therefore captured various mental disorders while also maintaining unity through a common core diagnosis among the targets. That said, the focus on BPD does not imply that BPD represents a prototypical form of mental disorder. Similarly, we restricted target photos to women to avoid potential gender-related biases related to mental disorder (e.g., Jones and Cochrane 1981). These restrictions limit the generalizability of the findings to women with a primary diagnosis of BPD, though they may extend to other mental disorders by virtue of their high comorbidity with other mental illnesses. Along these lines, we usually employed the term "mental disorder" throughout this work for its universality and to avoid the complex (but more specific) term "BPD," though specific ratings of BPD and MDD produced similar results as the more general umbrella term.

Finally, the studies relied on participants recruited online through MTurk. Despite requiring open-ended responses for some answers and including several attention check questions, we cannot confirm that individuals sampled online fully attended to or complied with the study's procedures as well as we can in the lab. Participants generally provided favorable comments, however, and many took the time to answer the open-ended questions and provide feedback at the completion of the study. Furthermore, data quality is generally consistent between laboratory and MTurk participants (Buhrmester et al. 2011; Goodman et al. 2013; Peer et al. 2014) and online participants provide the benefits of greater diversity in age, gender, race/ethnicity, and education compared to traditional laboratory samples that typically employ undergraduate students (Henrich et al. 2010).

Conclusion

Despite these limitations, the present findings may help to heighten awareness of the potential power of facial appearance in social and clinical contexts. Humans can extract an impressive amount of information from facial appearance without conscious awareness (e.g., Rule et al. 2008, 2010). The current findings provide evidence that this may extend to information about the presence of mental disorder, with perceptions of individuals' depression, emotional instability, anxiety, happiness, anger, disgust, and distress potentially influencing judgments of their mental health. Lay beliefs often hold that one can identify individuals with mental disorder through physical characteristics in their appearance. Future studies should therefore corroborate the accuracy of people's perceptions of mental disorder and the consequences that these categorizations may have for the stigmatization and well-being of individuals with mental disorder.

Acknowledgments The authors would like to thank the target individuals who allowed us to take their photograph for this research, as well as Sathaya Amirthavasagam, Gabriella Baumann, Chad Danyluck, Bryanna Graves, Jaeger Lam, Jessica Maxwell, Achala H. Rodrigo, Konstantin Tskhay, and especially Dr. Daniel E. Re for their comments and assistance. This study was supported in part by funding from the Social Sciences and Humanities Research Council of Canada to ARD and NOR, and from the Canadian Institutes of Health Research to ACR.

Appendix 1

Analyses of Participant Demographics

We tested for differences in accuracy according to participant demographics by aggregating the data from all of the studies and then regressing the participants' sensitivity correlations onto their age, gender, personal history of psychotherapy/counselling, and current use of psychotropic medications as fixed effects in a multilevel model in which we nested the scores within each study and participant (for the within-subjects design of Study 2B), estimating the random intercept. Prior to analysis, we dummy-coded the dichotomous variables and grand-mean centered all of the predictors. None of the variables produced significant main effects, all $|t|$s ≤ 1.35, ps $\geq .18$, though (unsurprisingly) both the random effects of study, $B = .0047$, $SE = .0002$, Wald $Z = 2.91$, $p = .004$, and of participant, $B = .0029$, $SE = .0007$, Wald $Z = 3.88$, $p < .001$, indicated significant variability. Thus, the participants did not vary according to their age, gender, personal experience with therapy, or present psychotropic medication use.

Analyses of Target Demographics

We also explored the relationship between the targets' demographic and clinical characteristics and the extent to which perceivers rated them as having a mental disorder. To do this, we averaged the continuous ratings of mental disorder for each target across perceivers within Study 1 and each of the conditions in Study 3 (all interrater reliabilities Cronbach's $\alpha \geq .89$) and measured the association between this consensus mental disorder score with targets' scores on the measures collected.

Most of the variables were not normally distributed. We therefore analyzed the relationships between them using nonparametric Spearman correlations of the rank-transformed variables. Results revealed that targets' ages significantly correlated with the legibility of their mental health in each of the samples, all rs $\geq .26$, all ps $\leq .04$, such that older targets were judged as more likely to have a mental disorder. However, none of the targets' depression (BDI), impulsivity (BIS), emotional stability (DERS), global functioning (GAF), intelligence (WTAR), or number of BPD symptoms significantly correlated with the mean consensus ratings of their likelihood of mental disorder in any of the studies, all $|r|$s $\leq .25$, all ps $\geq .07$.

Finally, we ran the same analyzes using the mean values for each target based on participants' ratings of BPD (Cronbach's $\alpha = .87$) and MDD (Cronbach's $\alpha = .94$) using Spearman correlations. Consistent with the above finding, older targets were more likely to be judged as having BPD or MDD (rs $> .28$, ps $< .03$). Higher ratings of BPD were also associated with self-reported depression (BDI; $r = .29$, $p = .03$), with all other correlations non-significant (all ps $\geq .06$).

Appendix 2

See Table 4.

Table 4 Demographic characteristics of the participants serving as perceivers for all studies

Condition	Mental disorder	Borderline personality disorder	Major depressive disorder	Multiple sclerosis	Depressed	Anxious	Impulsive	Emotional instability
Corresponding figure in main text	1A	1A	1A	1A	1A	1A	1A	1A
N	45	40	42	38	31	31	32	35
Gender M/F/T	22/23/0	17/23/0	26/16/0	12/26/0	17/14/0	16/15/0	16/16/0	15/20/0
Mean age	40.6	34.4	35.3	36.1	34.1	33.7	34.4	36.9
Age (*SD*)	11.9	13.1	12.6	13.6	12.5	11.1	11.0	12.8
Ethnicity								
Caucasian	80.0 %	72.5 %	71.4 %	78.9 %	64.5 %	70.0 %	87.5 %	62.9 %
Black	2.2 %	5.0 %	7.1 %	5.3 %	12.9 %	12.9 %	6.3 %	2.9 %
Native American	2.2 %	0 %	2.4 %	0 %	0 %	3.2 %	0 %	5.7 %
Hisp./Span./Mexican	2.2 %	15.0 %	4.8 %	5.3 %	12.9 %	6.5 %	3.1 %	17.1 %
Asian/other	0 %	7.5 %	16.7 %	10.5 %	9.7 %	9.7 %	3.1 %	11.4 %
Education level								
Below high school	0 %	0 %	0 %	0 %	0 %	0 %	6.3 %	0 %
High school	24.4 %	47.5 %	35.7 %	34.2 %	19.4 %	45.2 %	25.0 %	37.1 %
Associate's degree/diploma	24.4 %	15.0 %	19.0 %	15.8 %	25.8 %	9.7 %	28.1 %	25.7 %
Bachelor's degree	37.8 %	3.0 %	31.0 %	34.2 %	32.3 %	41.9 %	21.9 %	28.6 %
Master's degree	6.7 %	2.5 %	9.5 %	13.2 %	19.4 %	0 %	5.6 %	5.7 %
Doctoral degree	0 %	5.0 %	4.8 %	2.6 %	3.2 %	3.2 %	3.1 %	2.9 %
Employment status								
Unemployed	17.8 %	20.0 %	14.3 %	21.1 %	16.1 %	16.1 %	21.9 %	25.7 %
Employed full-time	64.4 %	32.5 %	45.2 %	58.9 %	58.1 %	41.9 %	43.8 %	34.3 %
Employed part-time	15.6 %	22.5 %	26.2 %	13.2 %	16.1 %	25.8 %	12.5 %	22.9 %
Full-time student	2.2 %	15.0 %	9.5 %	5.3 %	9.7 %	12.9 %	6.3 %	8.6 %
Part-time student	0 %	2.5 %	0 %	0 %	0 %	3.2 %	3.1 %	2.9 %
Disability/Assistance	0 %	7.5 %	9.5 %	2.6 %	0 %	0 %	12.5 %	5.7 %

Table 4 continued

Condition	Mental disorder	Borderline personality disorder	Major depressive disorder	Multiple sclerosis	Depressed	Anxious	Impulsive	Emotional instability
Mental health								
Self-reported mental illness	8	10	9	10	4	2	2	5
Currently taking psychotropic medication	8	13	9	5	5	6	7	7
Ever in counseling or psychotherapy	12	22	10	10	10	9	12	13

Condition	Introversion	Emotional expressions	Distressed	Inverted faces	Blurred faces	Elliptically-cropped faces	Upper face	Lower face
Corresponding figure in main text	1A	1A	1A	1B	1C	1D	1E	1F
N	32	43	37	30	33	43	52	52
Gender M/F/T	10/20/0	21/22/0	16/19/2	10/19/1	8/25/0	16/25/2	26/26/0	38/13/1
Mean age	36.7	35.9	34.5	34.0	36.5	36.1	35.2	39.8
Age (*SD*)	11.7	13.0	12.5	12.5	12.1	14.4	10.6	16.6
Ethnicity								
Caucasian	90.0 %	86.0 %	70.2 %	66.7 %	75.8 %	70.0 %	75.0 %	88.5 %
Black	0 %	7.0 %	13.5 %	10.0 %	3.0 %	9.3 %	5.8 %	3.8 %
Native American	3.3 %	0 %	0 %	10.0 %	0 %	0 %	0 %	0 %
Hisp./Span./Mexican	0 %	4.6 %	5.4 %	13.3 %	15.2 %	13.6 %	7.7 %	1.9 %
Asian/other	6.7 %	0 %	10.8 %	0 %	6.1 %	7.0 %	9.6 %	5.8 %
Education level								
Below high school	0 %	0 %	0 %	3.3 %	0 %	0 %	0 %	3.8 %
High school	23.3 %	34.9 %	29.7 %	43.3 %	18.2 %	41.9 %	38.5 %	32.7 %
Associate's degree/diploma	13.3 %	23.3 %	29.7 %	30.0 %	24.2 %	14.0 %	25.0 %	23.1 %
Bachelor's degree	46.7 %	23.3 %	18.9 %	16.7 %	33.3 %	34.9 %	30.8 %	36.5 %
Master's degree	10.0 %	14.0 %	13.5 %	6.7 %	21.2 %	7.0 %	5.8 %	1.9 %

Table 4 continued

Condition	Introversion	Emotional expressions	Distressed	Inverted faces	Blurred faces	Elliptically-cropped faces	Upper face	Lower face
Doctoral degree	6.7 %	4.7 %	8.1 %	0 %	3.0 %	2.3 %	0 %	1.9 %
Employment status								
Unemployed	23.3 %	16.3 %	13.5 %	16.7 %	15.2 %	30.2 %	25.0 %	30.8 %
Employed full-time	56.7 %	55.8 %	62.2 %	43.3 %	60.6 %	30.2 %	36.5 %	34.6 %
Employed part-time	16.7 %	14.0 %	10.8 %	20.0 %	15.2 %	20.9 %	23.1 %	26.9 %
Full-time student	0 %	11.6 %	8.1 %	10.0 %	9.1 %	14.0 %	11.5 %	0 %
Part-time student	0 %	0 %	0 %	0 %	0 %	0 %	1.9 %	7.7 %
Disability/Assistance	3.3 %	2.3 %	5.4 %	10.0 %	0 %	4.7 %	1.9 %	0 %
Mental health								
Self-reported mental illness	6	6	9	9	7	12	8	7
Currently taking psychotropic medication	6	7	8	7	8	11	7	4
Ever in counseling or psychotherapy	12	10	12	14	14	12	17	19

M male, *F* female, *T* transgender, *Hisp.* Hispanic, *Span.* Spanish

Appendix 3

See Table 5.

Table 5 Descriptive statistics and significance tests for target ratings of traits and emotional expressions in Studies 2A and 2B

	Raters	BPD		Healthy		t	p	r	α
	N	M	SD	M	SD				
Study 2A									
Anxiousness	31	3.32	(.88)	3.15	(.58)	1.08	.29	.20	.96
Depression	31	4.02	(1.02)	3.76	(.87)	1.42	.17	.25	.94
Emotional instability	35	4.11	(.70)	3.90	(.67)	1.77	.08	.30	.96
Impulsivity	32	4.01	(.72)	3.86	(.70)	1.18	.25	.21	.82
Introversion	30	3.70	(.84)	3.82	(.72)	−.78	.44	.14	.84
Study 2B									
Distressed	37	3.37	(.87)	3.01	(.81)	1.85	.07	.29	.96
Anger	43	2.15	(.62)	1.88	(.57)	1.75	.09	.26	.97
Disgust	43	2.11	(.53)	1.82	(.36)	2.50	.02	.36	.98
Fear	43	1.80	(.29)	1.70	(.26)	1.34	.18	.20	.96
Happiness	43	2.32	(.96)	2.77	(.99)	−1.79	.08	.27	.97
Sadness	43	2.23	(.55)	2.15	(.59)	.51	.61	.08	.98
Surprise	43	1.74	(.28)	1.71	(.30)	.35	.73	.05	.96

t-tests and accompanying statistics refer to independent-samples tests comparing participants' perceptions of targets from the two groups. Study 2B employed a within-subjects design for ratings of emotional expression

BPD borderline personality disorder, *M* mean, *SD* standard deviation, *N* sample size, α interrater reliability Cronbach's α

References

Akabaliev, V., Sivkov, S., Mantarkov, M., & Ahmed-Popova, F. (2011). Minor physical anomalies in patients with bipolar I disorder and normal controls. *Journal of Affective Disorders, 135*(1–3), 193–200. doi:10.1016/j.jad.2011.07.019.

Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology, 95*(4), 918–932.

Archinard, M., Haynal-Reymond, V., & Heller, M. (2000). Doctor's and patients' facial expressions and suicide reattempt risk assessment. *Journal of Psychiatric Research, 34*(3), 261–262. doi:10.1016/S0022-3956(00)00011-X.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for beck depression inventory-II*. San Antonio, TX: Psychological Corporation.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*(4), 645–657. doi:10.1037/0022-3514.62.4.645.

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*(3), 546–553.

Buckley, P. F., Dean, D., Bookstein, F. L., Han, S., Yerukhimovich, M., Min, K. J., & Singer, B. (2005). A three-dimensional morphometric study of craniofacial shape in schizophrenia. *American Journal of Psychiatry, 162*(3), 606–608. doi:10.1176/appi.ajp.162.3.606.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5. doi:10.1177/1745691610393980.

Carcone, D., Tokarz, V. L., & Ruocco, A. C. (2015). A systematic review on the reliability and validity of semi-structured diagnostic interviews for borderline personality disorder. *Canadian Psychology, 56*(2), 208–226.

Clark, L. A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103*, 103–116.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum and Associates.

Davison, K. (2006). Historical aspects of mood disorders. *Psychiatry, 5*(4), 115–118. doi:10.1383/psyt.2006.5.4.115.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 42*(3), 285–290.

Eaton, N. R., Krueger, R. F., Keyes, K. M., Skodol, A. E., Markon, K. E., Grant, B. F., & Hasin, S. H. (2011). Borderline personality disorder co-morbidity: Relationship to the internalizing-externalizing structure of common mental disorders. *Psychological Medicine, 41*(5), 1041–1050.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES–a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42*(1), 351–362. doi:10.3758/BRM.42.1.351.

Ekman, P. (1978). Facial signs: Facts, fantasies, and possibilities. In T. Sebok (Ed.), *Sight, sound, and sense* (pp. 124–156). Bloomington, I.N.: Indiana University Press.

Ekman, P., & Friesen, W. V. (1974). Nonverbal behavior and psychopathology. In R. J. Friedman & M. M. Katz (Eds.), *The psychology of depression: Contemporary theory and research* (pp. 203–232). Washington, D.C.: Winston & Sons.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. (2002). *Structured clinical interview for DSM-IV Axis I disorders—Patient edition (SCID-I/P, 11/2002 revision)*. New York, NY: Biometrics Research Department, New York Psychiatric Institute.

Fowler, K. A., Lilienfeld, S. O., & Patrick, C. J. (2009). Detecting psychopathy from thin slices of behavior. *Psychological Assessment, 21*(1), 68–78. doi:10.1037/a0014938.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89. doi:10.1146/annurev.clinpsy.1.102803.143810.

Gilman, S. L. (1982). *Seeing the insane*. New York: Wiley.

Glenn, C. R., & Klonsky, E. D. (2009). Emotion dysregulation as a core feature of borderline personality disorder. *Journal of Personality Disorders, 23*(1), 20–28. doi:10.1521/pedi.2009.23.1.20.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception, 34*(1), 77–86. doi:10.1068/p5370.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26*(3), 213–224. doi:10.1002/bdm.1753.

Grant, B. F., Chou, S. P., Goldstein, R. B., Huang, B., Stinson, F. S., Saha, T. D., & Ruan, W. J. (2008). Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of Clinical Psychiatry, 69*(4), 533–545.

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment, 26*(1), 41–54. doi:10.1023/B:Joba.0000007455.08539.94.

Gratz, K. L., Tull, M. T., Breetz, A. A., Matusiewicz, A. M., & Lejeuz, C. W. (2013). Multimodal examination of emotion regulation difficulties as a function of co-occurring avoidant personality disorder among women with borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment, 4*(4), 304–314.

Gunderson, J. G. (2010). Revising the borderline diagnosis for DSM-V: an alternative proposal. *Journal of Personality Disorders, 24*(6), 694–708. doi:10.1521/pedi.2010.24.6.694.

Hall, R. C. W. (1995). Global assessment of functioning: A modified scale. *Psychosomatics, 36*(3), 267–275.

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality, 42*(6), 1476–1489. doi:10.1016/J.Jrp.06.013.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. doi:10.1017/S0140525X0999152X.

Hess, U., Adams, R. B, Jr., Simard, A., Stevenson, M. T., & Kleck, R. E. (2012). Smiling and sad wrinkles: Age-related changes in the face and the perception of emotions and intentions. *Journal of Experimental Social Psychology, 48*, 1377–1380.

Holtzman, N. S. (2011). Facing a psychopath: Detecting the dark triad from neutral faces, using prototypes from the Personality Faceaurus. *Journal of Research in Personality, 45*(6), 648–654. doi:10.1016/j.jrp.2011.09.002.

Jane, J. S., Pagan, J. L., Turkheimer, E., Fiedler, E. R., & Oltmanns, T. F. (2006). The interrater reliability of the Structured Interview for DSM-IV Personality. *Comprehensive Psychiatry, 47*(5), 368–375. doi:10.1016/j.comppsych.2006.01.009.

Jones, L., & Cochrane, R. (1981). Stereotypes of mental illness: A test of the labeling hypothesis. *International Journal of Social Psychiatry, 27*(2), 99–107.

Jones, A. L., Kramer, R. S. S., & Ward, R. (2012). Signals of personality and health: The contributions of facial shape, skin texture, and viewing angle. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 1353–1361. doi:10.1037/a0027078.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 593–602.

Kleiman, S., & Rule, N. O. (2013). Detecting suicidality from facial appearance. *Social Psychological and Personality Science, 4*(4), 453–460. doi:10.1177/1948550612466115.

Klein, D. N., Ouimette, P. C., Kelly, H. S., Ferro, T., & Riso, L. P. (1994). Test-retest reliability of team consensus best-estimate diagnoses of axis I and II disorders in a family study. *American Journal of Psychiatry, 151*(7), 1043–1047.

Koenigsberg, H. W., Harvey, P. D., Mitropoulou, V., Schmeidler, J., New, A. S., Goodman, M., & Siever, L. J. (2002). Characterizing affective instability in borderline personality disorder. *American Journal of Psychiatry, 159*(5), 784–788. doi:10.1176/appi.ajp159.5.784.

Kramer, R. S. S., & Ward, R. (2010). Internal facial features are signals of personality and health. *The Quarterly Journal of Experimental Psychology, 63*(11), 2273–2287. doi:10.1080/17470211003770912.

Lane, A., Kinsella, A., Murphy, P., Byrne, M., Keenan, J., Colgan, K., & O'Callaghan, E. (1997). The anthropometric assessment of dysmorphic features in schizophrenia as an index of its developmental origins. *Psychological Medicine, 27*(5), 1155–1164. doi:10.1017/S0033291797005503.

Little, A. C., McPherson, J., Dennington, L., & Jones, B. C. (2011). Accuracy in assessment of self-reported stress and a measure of health from static facial information. *Personality and Individual Differences, 51*(6), 693–698. doi:10.1016/j.paid.2011.06.010.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates.

Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging, 2*, 64–69.

Manouilenko, I., Eriksson, J. M., Humble, M. B., & Bejerot, S. (2014). Minor physical anomalies in adults with autism spectrum disorder and healthy controls. *Autism Research and Treatment, 2014*, 743482. doi:10.1155/2014/743482.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255–260. doi:10.1016/S1364-6613(02)01903-4.

McGrath, J., El-Saadi, O., Grim, V., Cardy, S., Chapple, B., Chant, D., & Mowry, B. (2002). Minor physical anomalies and quantitative measures of the head and face in patients with psychosis. *Archives of General Psychiatry, 59*(5), 458–464. doi:10.1001/archpsyc.59.5.458.

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.

Oltmanns, T. F., Friedman, J. N. W., Fiedler, E. R., & Turkheimer, E. (2004). Perceptions of people with personality disorders based on thin slices of behavior. *Journal of Research in Personality, 38*(3), 216–229. doi:10.1016/S0092-6566(03)00066-7.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768–774. doi:10.1002/1097-4679(199511)51:6.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*(4), 1023–1031. doi:10.3758/s13428-013-0434-y.

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition, 24*(5), 490–524. doi:10.1521/soco.2006.24.5.607.

Pfohl, B., Blum, N., & Zimmerman, M. (1997). *Structured interview for DSM-IV personality*. Washington, DC: American Psychiatric Press.

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50-ms. *Journal of Experimental Social Psychology, 44*(4), 1100–1105. doi:10.1016/j.jesp.2007.12.001.

Rule, N. O., & Ambady, N. (2010). Democrats and republicans can be differentiated from their faces. *PLoS One, 5*, e8733. doi:10.1371/journal.pone.0008733.

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*(5), 1019–1028. doi:10.1037/a0013194.

Rule, N. O., Garrett, J. V., & Ambady, N. (2010). On the perception of religious group membership from faces. *PLoS One, 5*, e14241. doi:10.1371/journal.pone.0014241.

Ruocco, A. C., Lam, J., & McMain, S. F. (2014). Subjective cognitive complaints and functional disability in patients with borderline personality disorder and their nonaffected first-degree relatives. *Canadian Journal of Psychiatry, 59*(6), 335–344.

Scott, N. J., Kramer, R. S. S., Jones, A. L., & Ward, R. (2013). Facial cues to depressive symptoms and their associated personality attributions. *Psychiatry Research, 208*(1), 47–53. doi:10.1016/j.psychres.2013.02.027.

Sharp, C., Wright, A. G., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general ('g') and specific ('s') factors? *Journal of Abnormal Psychology, 124*, 387–398.

Slepian, M. L., Bogart, K. R., & Ambady, N. (2014). Thin-slice judgments in the clinical context. *Annual Review of Clinical Psychology, 10*, 131–153. doi:10.1146/annurev-clinpsy-090413-123522.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46*(2), 225–245.

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology, 117*(3), 647–661. doi:10.1037/a0012532.

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: the accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*(5), 1202–1216. doi:10.1037/a0013314.

Waddington, J. L., Lane, A., Larkin, C., & O'Callaghan, E. (1999). The neurodevelopmental basis of schizophrenia: Clinical clues from cerebro-craniofacial dysmorphogenesis, and the roots of a lifetime trajectory of disease. *Biological Psychiatry, 46*(1), 31–39.

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*(1), 120–128. doi:10.1037//0022-3514.57.1.120.

Watson, D., Clark, L. A., & Chmielewski, M. (2008). Structures of personality and their relevance to psychopathology: II. Further articulation of a comprehensive unified trait structure. *Journal of Personality, 76*(6), 1545–1586. doi:10.1111/j.1467-6494.2008.00531.x.

Waxer, P. (1976). Nonverbal cues for depth of depression: Set versus no set. *Journal of Consulting and Clinical Psychology, 44*(3), 493. doi:10.1037//0022-006x.44.3.493.

Waxer, P. (1977). Nonverbal cues for anxiety: Examination of emotional leakage. *Journal of Abnormal Psychology, 86*(3), 306–314. doi:10.1037//0021-843x.86.3.306.

Wechsler, D. (2002). *WTAR: Wechsler test of adult reading manual*. San Antonio, TX: The Psychological Corporation.

Wolff, G., Pathare, S., Craig, T., & Leff, J. (1996). Community knowledge of mental illness and reaction to mentally ill people. *British Journal of Psychiatry, 168*(2), 191–198.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141–145. doi:10.1037/h0027474.

Zanarini, M. C., & Frankenburg, F. R. (2007). The essential nature of borderline psychopathology. *Journal of Personality Disorders, 21*(5), 518–535. doi:10.1521/pedi.2007.21.5.518.

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?*. Boulder, CO: Westview Press.

Zimmerman, M., & Mattia, J. I. (1999). Axis I diagnostic comorbidity and borderline personality disorder. *Comprehensive Psychiatry, 40*(4), 245–252. doi:10.1016/S0010-440X(99)90123-2.