Article

PERCEPTION

Perception
2020, Vol. 49(3) 310–329

DOI: 10.1177/0301006620901671
journals.sagepub.com/home/pec
SAGE

The McGill Face Database: Validation and Insights Into the Recognition of Facial Expressions of Complex Mental States

Gunnar Schmidtmann
Eye & Vision Research Group, School of Health Professions, University of Plymouth, UK

Ben J. Jennings
Centre for Cognitive Neuroscience, Division of Psychology, Department of Life Sciences, Brunel University, London, UK

Dasha A. Sandra and Jordan Pollock
Department of Psychology, McGill University, Montreal, Quebec, Canada

Ian Gold
Departments of Philosophy and Psychiatry, McGill University, Montreal, Quebec, Canada

Abstract

Current databases of facial expressions represent only a small subset of expressions, usually the basic emotions (fear, disgust, surprise, happiness, sadness, and anger). To overcome these limitations, we introduce a database of pictures of facial expressions reflecting the richness of mental states. A total of 93 expressions of mental states were interpreted by two professional actors, and high-quality pictures were taken under controlled conditions in front and side view. The database was validated in two experiments. First, a four-alternative forced-choice paradigm was employed to test the ability to select a term associated with each expression. Second, the task was to locate each face within a 2-D space of valence and arousal. Results from both experiments demonstrate that subjects can reliably recognize a great diversity of emotional states from facial

Corresponding author:
Gunnar Schmidtmann, Eye & Vision Research Group, School of Health Professions, University of Plymouth, Plymouth, Devon, UK.
Email: gschmidtmann@googlemail.com

expressions. While subjects' performance was better for front view images, the advantage over the side view was not dramatic. This is the first demonstration of the high degree of accuracy human viewers exhibit when identifying complex mental states from only partially visible facial features. The McGill Face Database provides a wide range of facial expressions that can be linked to mental state terms and can be accurately characterized in terms of arousal and valence.

Keywords
faces, face database, emotions, mental states, theory of mind

Date Received: 3 May 2019; accepted: 17 December 2019

Introduction

Faces represent a special, very complex class of visual stimuli and have been extensively studied in a wide range of research areas. In particular, facial expressions are among the most important sources of information about the mental states of others. The capacity to make mental state inferences, whether from faces or other sources, is known as theory of mind (ToM), and it is widely agreed that this capacity is essential to human social behaviour. There is also substantial evidence that a ToM deficit may be associated with a variety of clinical conditions, notably autism (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) and schizophrenia (Bora, Yucel, & Pantelis, 2009; Brüne, 2005; Harrington, Siegert, & McClure, 2005; Sprong, Schothorst, Vos, Hox, & Van Engeland, 2007). Hence, the assessment of ToM is important for the exploration of social cognition in healthy individuals as well as in some patients. It may also be useful to measure a change in the social capacities of patients in psychotherapy. The "Reading the Mind in the Eyes" test (BaronCohen et al., 1997, 2001) is a common ToM test in which participants have to choose a mental state term that best characterizes the expression in a picture of someone's eyes. However, only a small proportion of possible mental states are tested, and the stimuli themselves are of inconsistent quality with respect to image resolution, luminance and perspective. Most other comparable databases of facial expressions of mental states typically include only a small subset of expressions, typically the basic emotions proposed by Paul Ekman (1992): fear, disgust, surprise, happiness, sadness, and anger) – the emotional expressions that are considered universal. However, multiple secondary emotions where two or more primary emotions are mixed (e.g., hatred being a mix of anger and disgust) are highly underrepresented in the databases available. One exception is the "Mind Reading" database (DVD; Baron-Cohen, Golan, Wheelwright, & Hill, 2004) that contains a much wider range of mental states. The Mind Reading DVD is computer-based platform developed to help individuals diagnosed along the autism spectrum to recognize facial expressions. It contains 412 mental state concepts, each assigned to 1 of 24 mental state classes. However, it is designed for commercial and clinical use and specifically targets patients with autism spectrum disorder and Asperger syndrome. A list of popular face stimuli databases is shown in Table 1. Most databases represent only a very small subset of emotions encountered in daily life and often in exaggerated form. To overcome these limitations, we have developed and validated a large new database of pictures of facial expressions – the McGill Face Database – that reflects some of the richness of human mental states. The database contains high-resolution pictures of 93 expressions of mental states that were interpreted by two professional actors (one male and one female) in front

Table 1. Summary of Face Databases.

Database	Reference	No. images	Expressions
The Yale Face Database	Belhumeur, Hespanha, and Kriegman (1996)	165	Happy, sad, winking, sleepy, surprised
AR Face Database	Martinez and Benavente (1998)	3,000	n.s.
Karolinska-Directed Emotional Faces (KDEF)	Lundqvist, Flykt, and Ohman (1998)	4,900	Anger, happiness, surprise, disgust, sadness, fear, neutral
	Goeleven, Raedt, Leyman, and Verschuere (2008)	490	Angry, fearful, disgusted, happy, sad, surprised
Japanese Female Facial Expression (JAFFE)	Lyons, Akamatsu, Kamachi, and Gyoba (1998)	219	Anger, happiness, surprise, disgust, sadness, fear, neutral
Yale Face Database B+	Georghiades, Belhumeur, and Kriegman (2000)	4,050	n.s.
Palermo & Coltheart Faces	Palermo and Coltheart (2004)	336	Anger disgust, fear, happiness, neutrality, sadness, surprise
MMI (M & M Initiative)	Pantic, Valstar, Rademaker, and Maat (2005)	1,588	79, n.s.
BU-3DFE Database	Yin, Wei, Sun, Wang, and Rosato (2006)	2,500	Anger, disgust, fear, happy, sad, surprise, neutral
The Bosphorus database	Alyüz et al. (2008)	4,666	n.s.
Multi-PIE (Pose, Illumination, Expression)	Gross, Matthews, Cohn, Kanade, and Baker (2010)	750,000+	Neutral, smile, surprise, squint, disgust, scream
Genki-4K	Whitehill, Littlewort, Fasel, Bartlett, and Movellan (2009)	63,000	Smiling or nonsmiling
The MUG Face Database	Aifanti, Papachristou, and Delopoulos (2010)	70,645	Anger, fear, happiness, sadness, surprise
FACES	Ebner, Riediger, and Lindenberger (2010)	2,052	Neutral, sadness, disgust, fear, anger, happiness
Radboud Faces	Langner et al. (2010)	5,880	Angry, contemptuous, disgusted, fearful, happy, sad, surprised, neutral
Cohn-Kanade CK+	Lucey et al. (2010)	593 recordings, 10,708 frames	Anger, contempt, disgust, fear, happy, sadness, surprise
Indian Movie Face database (IMFDB)	Setty et al. (2013)	34,512	Anger, happiness, surprise, disgust, sadness, fear
DynEmo	Tcherkassof et al. (2013)	358 videos	n.s.
KinectFaceDB	Min, Kose, and Dugelay (2014)	156 images, 52 videos	Neutral, smile

Note. n.s. = not specified.

Table 2. Summary of Terms in the McGill Face Database.

	English		English
1	Accusing	48	Grateful
2	Affectionate	49	Guilty
3	Aghast	50	Hateful
4	Alarmed	51	Hopeful
5	Amused	52	Horrified
6	Annoyed	53	Hostile
7	Anticipating	54	Impatient
8	Anxious	55	Imploring
9	Apologetic	56	Incredulous
10	Arrogant	57	Indecisive
11	Ashamed	58	Indifferent
12	Assertive	59	Insisting
13	Baffled	60	Insulting
14	Bewildered	61	Interested
15	Cautious	62	Intrigued
16	Comforting	63	Irritated
17	Concerned	64	Jealous
18	Confident	65	Joking
19	Confused	66	Nervous
20	Contemplative	67	Offended
21	Contented	68	Panicked
22	Convinced	69	Pensive
23	Curious	70	Perplexed
24	Deciding	71	Playful
25	Decisive	72	Preoccupied
26	Defiant	73	Puzzled
27	Depressed	74	Reassuring
28	Desire	75	Reflective
29	Despondent	76	Regretful
30	Disappointed	77	Relaxed
31	Dispirited	78	Relieved
32	Distrustful	79	Resentful
33	Dominant	80	Sarcastic
34	Doubtful	81	Satisfied
35	Dubious	82	Serious
36	Eager	83	Skeptical
37	Earnest	84	Stern
38	Embarrassed	85	Suspicious
39	Encouraging	86	Sympathetic
40	Entertained	87	Tentative
41	Enthused	88	Terrified
42	Fantasizing	89	Thoughtful
43	Fascinated	90	Threatening
44	Fearful	91	Uneasy
45	Flirtatious	92	Upset
46	Flustered	93	Worried
47	Friendly		

and side view – 372 images in total. In this article, we present two different experiments to investigate subjects' ability to recognize the facial expressions in the database. In Experiment 1, we employ a four-alternative forced-choice (4-AFC) paradigm, based on previous studies (BaronCohen et al., 1997, 2001). The task for the observer in this experiment was to choose, out of four terms, the one that best identifies the mental state expressed. Given that a particular "correct" term is only a representation of the actors' interpretations of the mental state, a second validation experiment (Experiment 2) was carried out, which did not rely on the semantics of the mental state terms. Instead, the observers located each face within a two-dimensional (2-D) space of valence and arousal (mental state – space) employing a "point-and-click" paradigm (Jennings, Yu, & Kingdom, 2017).

Database

Actor Recruitment

Five male and five female professional native English-speaking actors were invited to take part in an audition. The actors' performance was judged by a panel of two of the authors and a theatre-experienced Professor of Drama and Theatre in the McGill Department of English. During the audition, one male and one female actor engaged in various improvisation exercises. The "best actors" were those who exhibited the most precise, nuanced, and yet readable range of emotional expression in their faces, that is, that clarity of emotional expression – as captured by the camera – was paramount. Some actors were better able to convey different emotions through subtle recalibration of facial expression, while others either got "stuck in look" or fell into exaggerated or melodramatic countenances. The two best performing actors (male, age 29; female, age 23) were chosen to take part in a photo shoot based on a majority vote. The actors gave informed consent and signed an agreement allowing for the pictures to be used for research and other noncommercial purposes. The actors were compensated for their work.

Images

Equipment. The pictures were taken by a professional photographer with a Canon 70 D digital camera mounted on a tripod at a distance of 1.5 m from the actor. The optic was a Canon 85 mm, *f*1.8 with a shutter speed of $\frac{1}{60}$ and an aperture of *f*5.6 and a sensitivity of ISO 100. Two separate flashes, a Canon 580 EX and a Canon 430 EXII (both set with exposure compensation at +1), were placed at the appropriate distance. One of the flashes had a reflector umbrella.

Image acquisition. The pictures were taken in two separate sessions at a studio specifically prepared for that purpose. During the sessions, the actor was positioned in front of a white screen. The instructor provided the mental state term and read the corresponding short explication provided in the Glossary in Appendix B of Baron-Cohen et al. (2001). The actor was given as much time as needed to prepare the interpretation for the relevant expression. When the actor gave a hand signal to the photographer, a single picture was taken in front view. Importantly, to guarantee a natural interpretation of a given expression, we did not restrict the head tilt. The actor then immediately turned to face a mark 30° from the camera, and a second picture was taken. This procedure was repeated three to four times for each of 93 mental state terms used in the Reading the Mind in the Eyes test (Baron-Cohen et al., 2001; Table 2).

Table 3. Percent Correct for the Images Averaged Across 32 Subjects.

		Male		Female				Male		Female	
		Front	Side	Front	Side			Front	Side	Front	Side
1	Accusing	36.36*	60.61	54.55	30.3*	48	Grateful	54.55	36.36*	51.52	81.82
2	Affectionate	30.3*	18.18*	69.7	63.64	49	Guilty	24.24*	24.24*	24.24*	18.18*
3	Aghast	78.79	78.79	84.85	87.88	50	Hateful	60.61	57.58	60.61	63.64
4	Alarmed	66.67	54.55	84.85	57.58	51	Hopeful	48.48	30.3*	57.58	63.64
5	Amused	69.7	51.52	84.85	78.79	52	Horrified	78.79	93.94	45.45	27.27*
6	Annoyed	72.73	51.52	69.7	81.82	53	Hostile	51.52	30.3*	39.39	72.73
7	Anticipating	30.3*	48.48	36.36*	78.79	54	Impatient	36.36*	72.73	51.52	66.67
8	Anxious	48.48	66.67	54.55	60.61	55	Imploring	27.27	48.48	66.67	39.39
9	Apologetic	12.12*	18.18*	57.58	45.45	56	Incredulous	48.48	57.58	54.55	45.45
10	Arrogant	72.73	18.18*	48.48	36.36*	57	Indecisive	45.45	63.64	63.64	51.52
11	Ashamed	18.18*	12.12*	42.42	42.42	58	Indifferent	39.39	48.48	57.58	66.67
12	Assertive	60.61	51.52	57.58	21.21	59	Insisting	63.64	60.61	51.52	45.45
13	Baffled	42.42	36.36*	63.64	60.61	60	Insulting	60.61	36.36*	60.61	30.3*
14	Bewildered	63.64	84.85	63.64	81.82	61	Interested	30.3*	36.36*	24.24	30.3*
15	Cautious	51.52	54.55	54.55	33.33*	62	Intrigued	36.36	60.61	45.45	75.76
16	Comforting	3.03*	12.12*	69.7	78.79	63	Irritated	48.48	39.39	63.64	69.7
17	Concerned	57.58	60.61	66.67	69.7	64	Jealous	36.36*	36.36*	15.15*	30.3*
18	Confident	21.21*	21.21*	84.85	51.52	65	Joking	75.76	78.79	72.73	63.64
19	Confused	51.52	60.61	54.55	81.82	66	Nervous	69.7	42.42	45.45	24.24
20	Contemplative	84.85	72.73	54.55	57.58	67	Offended	72.73	39.39	60.61	87.88
21	Contented	39.39	54.55	87.88	72.73	68	Panicked	84.85	78.79	54.55	78.79
22	Convinced	27.27	9.09*	39.39	27.27	69	Pensive	84.85	30.3*	72.73	63.64
23	Curious	27.27*	63.64	45.45	57.58	70	Perplexed	66.67	75.76	75.76	84.85
24	Deciding	66.67	75.76	48.48	51.52	71	Playful	90.91	90.91	90.91	72.73
25	Decisive	45.45	42.42	18.18*	18.18*	72	Preoccupied	15.15	42.42	42.42	60.61
26	Defiant	66.67	63.64	42.42	27.27*	73	Puzzled	57.58	72.73	81.82	84.85
27	Depressed	45.45	33.33*	69.7	54.55	74	Reassuring	27.27*	21.21*	51.52	57.58
28	Desire	21.21*	33.33*	63.64	72.73	75	Reflective	60.61	72.73	18.18*	39.39
29	Despondent	60.61	39.39	54.55	60.61	76	Regretful	27.27*	33.33*	33.33*	54.55
30	Disappointed	66.67	27.27*	78.79	75.76	77	Relaxed	42.42	39.39	87.88	69.7
31	Dispirited	54.55	51.52	75.76	87.88	78	Relieved	27.27*	39.39	36.36*	60.61
32	Distrustful	81.82	48.48	60.61	54.55	79	Resentful	54.55	30.3*	33.33*	30.3*
33	Dominant	78.79	45.45	60.61	54.55	80	Sarcastic	51.52	54.55	51.52	66.67
34	Doubtful	81.82	54.55	78.79	63.64	81	Satisfied	81.82	51.52	66.67	63.64
35	Dubious	57.58	39.39	57.58	54.55	82	Skeptical	51.52	57.58	66.67	69.7
36	Eager	72.73	87.88	66.67	45.45	83	Serious	72.73	72.73	57.58	69.7
37	Earnest	30.3*	33.33*	36.36*	33.33*	84	Stern	78.79	66.67	84.85	42.42
38	Embarrassed	36.36*	42.42	60.61	63.64	85	Suspicious	75.76	63.64	66.67	63.64
39	Encouraging	60.61	72.73	42.42	78.79	86	Sympathetic	15.15*	27.27*	51.52	57.58
40	Entertained	90.91	75.76	66.67	57.58	87	Tentative	57.58	36.36*	21.21*	63.64
41	Enthused	93.94	51.52	87.88	78.79	88	Terrified	81.82	81.82	84.85	90.91
42	Fantasizing	75.76	60.61	48.48	39.39	89	Thoughtful	60.61	90.91	39.39	48.48
43	Fascinated	66.67	66.67	57.58	69.7	90	Threatening	72.73	81.82	30.3*	30.3*
44	Fearful	72.73	60.61	69.7	69.7	91	Uneasy	66.67	72.73	63.64	69.7
45	Flirtatious	51.52	60.61	66.67	87.88	92	Upset	24.24*	27.27*	63.64	75.76
46	Flustered	66.67	63.64	63.64	60.61	93	Worried	78.79	57.58	78.79	69.7
47	Friendly	57.58	81.82	87.88	72.73						

Note. The guess rate is 25%. Asterisks indicate performances that are statistically not better than chance (χ^2 – Yates correction for continuity; $\alpha > .05$).

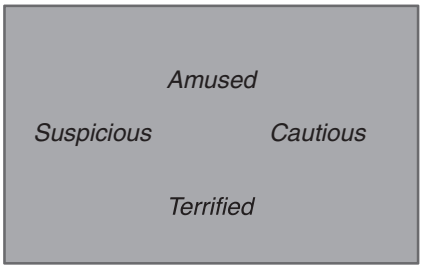

Figure 1. Experiment 1: Example of decision display. All 372 images in the database were presented in a random order in a single experimental block. Within one experimental trial, a picture was shown for 1 s, followed by a search display illustrated in the figure.

Image selection. A focus group, consisting of six referees (four females and two males) were presented with the different images for a given expression and asked to compare their quality and expressivity of mental state. Four out of six referees had to agree on a picture for it to be selected for inclusion in the database. The full database can be downloaded at: http://www.gunnar-schmidtmann.com/stimuli-software#McGillFaceDatabase.

Image specificities. The database contains 372 jpeg image files with a resolution of $5{,}472 \times 3{,}648$ pixel (colour space profile: sRGB IEC61966-2.1). The size of each image is 7.3 MB. The image files have not been postprocessed. Raw image files are available upon request from the first author.

Experiment 1

Methods

Subjects. All participants were recruited via the McGill Psychology Human Participant Pool or via public advertisements. Thirty-three individuals (7 males, 26 females; mean age 21 years, ±2.96 SD) participated in Experiment 1. All subjects were native English speakers and were naïve as to the purpose of the study. Subjects had normal or corrected-to-normal visual acuity. Informed consent was obtained from each observer. All experiments were approved by the McGill University Ethics committee and were conducted in accordance with the original Declaration of Helsinki.

Apparatus. The face stimuli were presented using MATLAB (MathWorks, Natick, Massachusetts, USA) (MATLAB R 2016b, MathWorks) on either a CRT monitor running with a resolution of $1{,}600 \times 1{,}200$ pixel and a frame rate of 60 Hz (mean luminance $40\frac{cd}{m^2}$) under the control of a PC (3.2 GHz) or on a MacBook Pro (2015, 3.1 GHz) with a monitor resolution of $2{,}560 \times 1{,}600$ pixel. The viewing distance was adjusted to guarantee an equal image size of $20.91^\circ \times 13.95^\circ$ on both systems. Experiments were performed in a dimly illuminated room. Routines from the Psychtoolbox-3 were employed to present the stimuli (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

Procedure. A 4-AFC paradigm was employed to test the ability of participants to correctly select the term associated with each picture in the database. All 372 pictures (93 male front view, 93 male side view, 93 female front view, and 93 female side view) were tested in one

experimental block. The images were presented in random order, different for every observer. Stimuli were presented for 1 s. This presentation time was based on previous results, where identification accuracy for the same face stimuli was measured as a function of presentation time (Schmidtmann, Sleiman, Pollack, & Gold, 2016). The presentation of the face image was followed by the presentation of the target (correct) term as well as three distractor terms. Importantly, to minimize a decision bias caused by specific terms, the distractor terms were randomly selected from the remaining 92 terms shown in Table 2. In other words, each observer was presented with different distractor terms for each face. The terms were presented on a midgrey screen in a diamond-like arrangement (see Figure 1), corresponding to the cursor keys on a computer keyboard, which were used to by the observers to make their choice. The target term could occur in one out of four locations, which was randomly determined. The task for the observer was to choose the term most appropriate to the expression in the picture. Participants were given a break after 93 presentations, that is, three breaks in total.

Results

Table 3 summarizes the performance (percent correct) across 33 subjects. The guess rate in a 4-AFC paradigm is 25%. χ^2 tests with a Yates (1934) correction for continuity ($p > .05$) were performed to determine whether performances were significantly different from chance level for a given term. Performances not significantly better than chance are shown by the by the asterisks in Table 3 and by the lines in Figures 2 and 3 showing the sorted percent correct performances for the actors in front and side view as bar plots. Results show that for the pictures of the female actor, subjects performed significantly better than chance in 78 of 93 images (84%) for the front view condition and 74 of 93 images (80%) of the side view pictures. For the male actor, subjects performed significantly better than chance in 67 of 93 images (72%) in front view and 61 of 93 images (66%) in side view. The nonsignificant terms are summarized in Table 4. Interestingly, 13 of these 52 nonsignificant cases occur in judgements of both the female and male actor. Furthermore, in 8 of these 52 terms, subjects performed no better than chance for 3 or 4 of the images. These terms are indicated by the the asterisks in Table 4.

In addition, we conducted parametric Pearson correlation between each combination of the stimuli tested in Experiment 1. Results show statistically significant correlations between results for the female faces in front and side view ($r = .555$, $p < .001$, $n = 93$), male faces in front and side view ($r = .598$, $p < .001$, $n = 93$), and female and male faces in front view ($r = .336$, $p = .001$, $n = 93$). All other correlations are presented in Table A1.

Experiment 2

Methods

Subjects. Thirty-two subjects participated in Experiment 2 (10 males, 22 females; mean age 22 years, ±4.13 SD).

Procedure. We employed a "point-and-click" task that did not rely on any semantic information being presented to observers during trials (Jennings et al., 2017). The complete set of images (372) was presented in a random order. Each image was displayed for 1 s followed by the 2-D mental state-space (Russell, 1980), presented until the observer submitted a response (Figure 4 shows the 2-D space). Once the 2-D space was displayed, the observers' task was

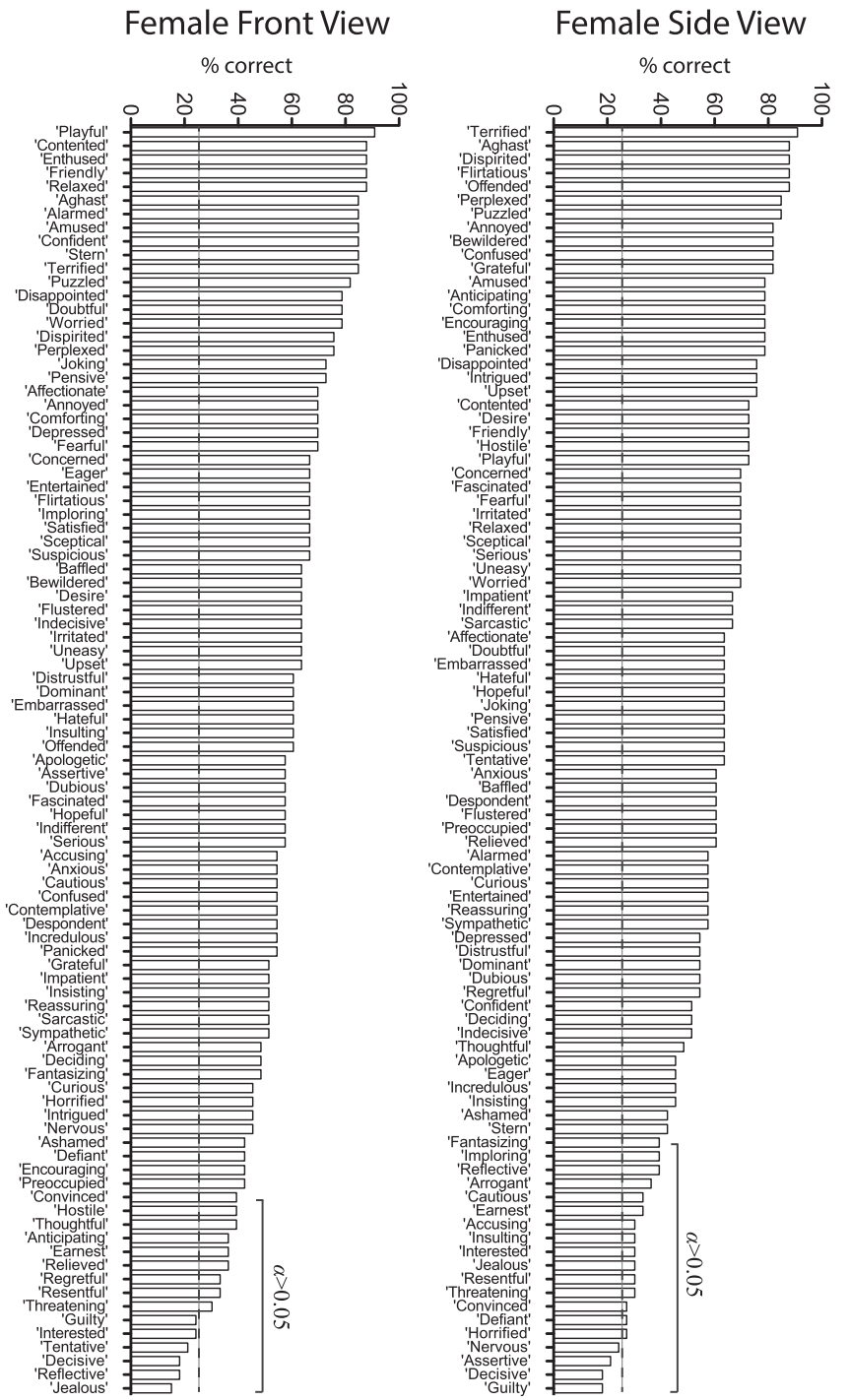

Figure 2. Bar plots showing percent correct for the 93 terms in the database for the female actor in both views. The dashed line represents the guessing rate (25%). Performances that are statistically not better than chance (χ^2 – Yates correction for continuity; $\alpha > .05$) are indicated by the solid lines in each graph.

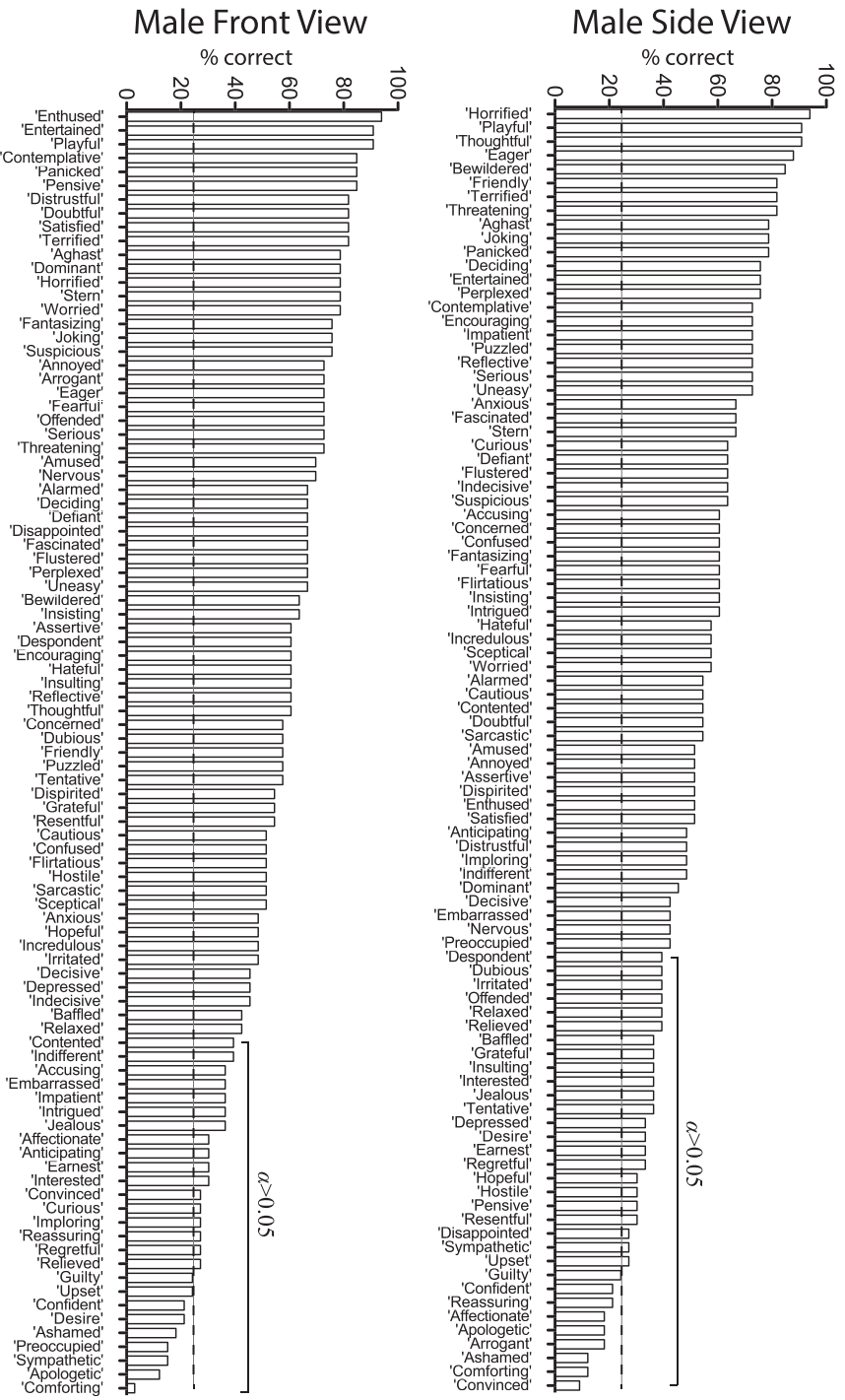

Figure 3. Bar plots showing percent correct for the 93 terms in the database for the male actor in both views. The dashed line represents the guessing rate (25%). Performances that are statistically not better than chance (χ^2 – Yates correction for continuity; $\alpha > .05$) are indicated by the solid lines in each graph.

Table 4. Summary of Terms (Sorted Alphabetically) in Which Participants' Performances Were Not Significantly Better Than Chance.

		Female		Male				Female		Male	
		Front	Side	Front	Side			Front	Side	Front	Side
1	Accusing		30.3	36.36		27	Hopeful				30.3
2	Affectionate			30.3	18.18	28	Horrified		27.27		
3	Anticipating	36.36		30.3		29	Hostile	39.39			30.3
4	Apologetic			12.12	18.18	30	Impatient			36.36	
5	Arrogant		36.36		18.18	31	Imploring		39.39	27.27	
6	Ashamed			18.18	12.12	32	Indifferent			39.39	
7	Assertive		21.21			33	Insulting		30.3		36.36
8	Baffled				36.36	34	Interested*	24.24	30.3	30.3	36.36
9	Cautious		33.33			35	Intrigued			36.36	
10	Comforting			3.03	12.12	36	Irritated				39.39
11	Confident			21.21	21.21	37	Jealous*	15.15		36.36	36.36
12	Contented			39.39		38	Nervous		24.24		
13	Convinced*	39.39	27.27	27.27	9.09	39	Offended				39.39
14	Curious			27.27		40	Pensive				30.3
15	Decisive	18.18	18.18			41	Preoccupied			15.15	
16	Defiant		27.27			42	Reassuring			27.27	21.21
17	Depressed				33.33	43	Reflective	18.18	39.39		
18	Desire			21.21	33.33	44	Regretful*	33.33		27.27	33.33
19	Despondent				39.39	45	Relaxed				39.39
20	Disappointed				27.27	46	Relieved*	36.36		27.27	39.39
21	Dubious				39.39	47	Resentful*	33.33	30.3		30.3
22	Earnest*	36.36	33.33	30.3	33.33	48	Sympathetic			15.15	27.27
23	Embarrassed			36.36		49	Tentative	21.21			36.36
24	Fantasizing		39.39			50	Thoughtful	39.39			
25	Grateful				36.36	51	Threatening	30.3	30.3		
26	Guilty*	24.24	18.18	24.24	24.24	52	Upset			24.24	27.27

Note. Asterisks indicate cases that were not significant in three or more conditions.

to click a computer mouse on the point within the space deemed most appropriate to the facial expression displayed in the image. The horizontal direction represented a rating of valence (pleasant vs. unpleasant) and the vertical direction a rating of arousal (low vs. high). Example emotions corresponding to different regions of the space are illustrated by the red text (not visible during testing) in Figure 4. The axes as well as the example mental states (red) were used to instruct the observer during training. To evaluate whether participants tended to locate facial expressions in similar regions of the 2-D space, we calculated an agreement score ($\eta_{agreement}$) for each image among 32 observers in the following way.

First, the median arousal (A_{median}) and valence (V_{median}) coordinates were calculated across all observer responses for a given condition. Second, the Euclidian distance (r) for each of the observers' response, and hence the mean (r_{mean}, see Equation (1)), was determined. Finally, these values were normalized (based on the highest mean value, r_{max}) and shifted according to the lowest value (r_{min}, see Equation (3)). This transformation produced agreement scores ($\eta_{agreement}$) so that a score of 1 corresponds to the greatest agreement between subjects and as the scores decrease, the agreement between subjects' decreases, that is, emotion ratings were less tightly clustered around the mean location (see

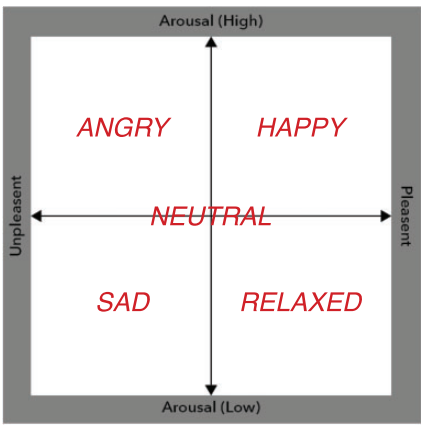

Figure 4. Experiment 2: The image was presented for 1 s, followed by the presentation of a valence-arousal space, extending from low to high arousal in one dimension and pleasant to unpleasant in the other dimension. Note: The red terms provide illustrations of the appropriate location of mental state terms used (the red text was not visible during testing).

Equation (3)). Figure 5 illustrates the procedure for four hypothetical data points located within a subsection of the arousal-valence space.

$$r_{mean} = \frac{1}{n}\sum_{1}^{i=n}\sqrt{(V_{median} - V_i)^2 + (A_{median} - A_i)^2} \tag{1}$$

$$r_{mean} = \frac{1}{n}\sum_{1}^{i=n} r_i \tag{2}$$

$$\eta_{agreement} = 1 - \frac{r_{mean}}{r_{max}} + r_{min} \tag{3}$$

Results

Tables 5 and 6 summarize the agreement scores ($\eta_{agreement}$), for the female and male face stimuli, respectively. To visualize the magnitude of the agreement scores within the valence-arousal space, three examples are illustrated in Figure 6. The circles are rendered with a radius equal to the values produced by Equation (1), and the corresponding agreement values are stated for comparison. The results for each of the 93 terms can be found in Supplemental Material Document 1. A Pearson correlation analysis between all stimulus types (female front view, female side view, male front view, and male side view) revealed statistically significant correlations between the following three conditions: (a) male front view versus female front view ($r = .34$, $p < .001$, $n = 93$), (b) male front view versus male side view ($r = .54$, $p < .001$, $n = 93$) and (c) female front view versus female side view ($r = .52$, $p < .001$, $n = 93$). Interestingly, there was no statistically significant correlation between female side view and male side view ($r = .074$, $p = .48$, $n = 93$).

In addition, we have created a method to visualize the amount of overlap between the different circular regions (i.e., r_{mean}, calculated according according to Equation (1)). The figures presented in Supplemental Material (Document 2) show matrices that show

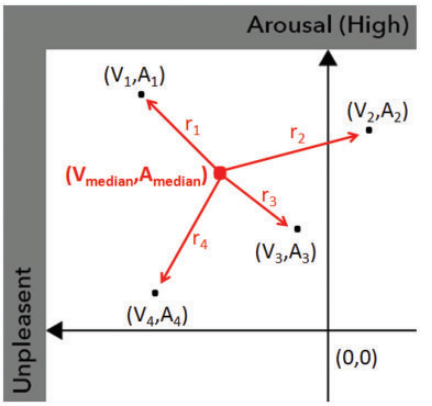

Figure 5. A subsection of the valence-arousal space showing four hypothetical responses (black dots); the red dot represents the mean valence and arousal. The agreement score ($\eta_{agreement}$) is determined by the mean Euclidian distance *r*.

combinations with no overlap with respect to r_{mean} in the valence-arounal space (shown in black). In addition, we have also calculated the overall proportion of combinations that have no overlap. This analysis revealed the following results: female front view: 17.8%; female side view: 15.9%; male front view: 13.3% and male side view: 7%. These matrices are very detailed and should be viewed in a higher magnification. The matrices can be found in Supplemental Material Document 2.

Relationships Between Experiment 1 and Experiment 2

A careful inspection of the raw data provided in the Supplemental Material (Document 1) reveals clear differences between the observers' decisions for the female and male interpretations of the mental states. Specifically, subjects selected coordinates in the left, unpleasant, more negative quadrants for one stimulus type (e.g., male) and the opposite, pleasant quadrants for the other stimulus type (e.g., female), or vice versa. This seems to suggest that specific facial expression can represent various mental states or that specific mental states can be interpreted (by the actor) in different, opposite ways. For instance, in the case of sarcastic, the subjects consistently selected the two quadrants in the unpleasant region (left: red and black data points) for the female version, but the pleasant quadrants for the male face, which is presumably based on the more positive interpretation of this particular mental state by the male actor (i.e., smiling). Interestingly, these differences are also reflected in the results for Experiment 1 for these terms, which are summarized in the Table 7. The differences between subjects' performances for male and female interpretations of mental states in Experiment 1 are particularly dramatic for anticipating (side view), comforting, confident, contented, and decisive. With the exception of decisive, where the performance is better for the male version, it is usually the female stimulus that elicits better performances. We believe that this is related to not only the actors' ability to express mental states, but also their interpretation of the mental state. Sarcastic, however, is an interesting case. Performances between the two stimulus types (female and male) are very similar in Experiment 1, but lead to completely different decisions in Experiment 2.

In a final analysis, parametric Pearson correlation tests were conducted between the percent correct performance for each stimulus in Experiment 1 and the agreement score $\eta_{agreement}$ for each stimulus in Experiment 2. This analysis showed statically significant

Table 5. Agreement Scores ($\eta_{agreement}$) for the Female in Front and Side View.

	Female								
	Front		Side			Front		Side	
1	Concerned	0.99	Enthused	0.982	48	Arrogant	0.834	Relaxed	0.844
2	Relieved	0.98	Eager	0.969	49	Resentful	0.834	Ashamed	0.844
3	Amused	0.966	Upset	0.961	50	Despondent	0.831	Sceptical	0.842
4	Playful	0.964	Relieved	0.96	51	Annoyed	0.829	Distrustful	0.84
5	Cautious	0.945	Guilty	0.954	52	Jealous	0.829	Resentful	0.84
6	Satisfied	0.939	Tentative	0.945	53	Joking	0.827	Contemplative	0.838
7	Friendly	0.934	Reassuring	0.943	54	Anxious	0.826	Reflective	0.836
8	Indecisive	0.925	Jealous	0.941	55	Thoughtful	0.821	Nervous	0.834
9	Accusing	0.921	Amused	0.941	56	Hopeful	0.819	Comforting	0.833
10	Relaxed	0.916	Playful	0.937	57	Puzzled	0.815	Incredulous	0.827
11	Confident	0.915	Impatient	0.933	58	Stern	0.813	Sympathetic	0.825
12	Fantasizing	0.912	Disappointed	0.905	59	Intrigued	0.812	Deciding	0.824
13	Comforting	0.909	Concerned	0.903	60	Reflective	0.811	Indifferent	0.823
14	Encouraging	0.897	Pensive	0.901	61	Indifferent	0.81	Encouraging	0.818
15	Reassuring	0.897	Cautious	0.9	62	Eager	0.802	Threatening	0.815
16	Tentative	0.896	Depressed	0.896	63	Earnest	0.8	Intrigued	0.815
17	Interested	0.891	Friendly	0.895	64	Guilty	0.8	Decisive	0.815
18	Assertive	0.89	Dubious	0.893	65	Threatening	0.799	Fascinated	0.805
19	Upset	0.886	Preoccupied	0.891	66	Desire	0.793	Desire	0.804
20	Defiant	0.885	Indecisive	0.891	67	Serious	0.78	Affectionate	0.803
21	Sarcastic	0.884	Regretful	0.89	68	Convinced	0.778	Hopeful	0.802
22	Regretful	0.883	Joking	0.888	69	Anticipating	0.777	Worried	0.799
23	Ashamed	0.881	Puzzled	0.887	70	Sympathetic	0.777	Fantasizing	0.798
24	Contented	0.879	Anxious	0.881	71	Imploring	0.776	Assertive	0.786
25	Disappointed	0.872	Flustered	0.876	72	Dominant	0.763	Bewildered	0.786
26	Entertained	0.871	Alarmed	0.876	73	Baffled	0.761	Entertained	0.783
27	Pensive	0.871	Suspicious	0.874	74	Insisting	0.76	Curious	0.782
28	Depressed	0.871	Flirtatious	0.874	75	Fascinated	0.759	Contented	0.782
29	Dispirited	0.87	Thoughtful	0.873	76	Incredulous	0.757	Embarrassed	0.775
30	Hostile	0.869	Imploring	0.869	77	Embarrassed	0.756	Sarcastic	0.775
31	Contemplative	0.866	Earnest	0.869	78	Affectionate	0.755	Confident	0.771
32	Irritated	0.865	Insisting	0.868	79	Insulting	0.742	Serious	0.771
33	Flirtatious	0.862	Convinced	0.867	80	Uneasy	0.735	Fearful	0.764
34	Preoccupied	0.862	Hostile	0.866	81	Horrified	0.734	Accusing	0.764
35	Enthused	0.86	Uneasy	0.865	82	Perplexed	0.724	Defiant	0.758
36	Decisive	0.86	Perplexed	0.864	83	Fearful	0.718	Annoyed	0.74
37	Nervous	0.857	Baffled	0.86	84	Bewildered	0.711	Insulting	0.735
38	Impatient	0.856	Interested	0.86	85	Dubious	0.707	Stern	0.734
39	Apologetic	0.855	Doubtful	0.859	86	Hateful	0.704	Anticipating	0.734
40	Confused	0.854	Grateful	0.853	87	Worried	0.7	Horrified	0.732
41	Distrustful	0.846	Irritated	0.85	88	Sceptical	0.699	Apologetic	0.727
42	Flustered	0.842	Confused	0.849	89	Alarmed	0.655	Hateful	0.712
43	Curious	0.841	Satisfied	0.849	90	Aghast	0.651	Panicked	0.679
44	Grateful	0.841	Dominant	0.847	91	Panicked	0.649	Terrified	0.672
45	Deciding	0.839	Dispirited	0.847	92	Terrified	0.57	Offended	0.66
46	Suspicious	0.835	Arrogant	0.846	93	Offended	0.532	Aghast	0.648
47	Doubtful	0.834	Despondent	0.846					

Note. Terms are sorted from high to low scores in each view.

Table 6. Agreement Scores ($\eta_{agreement}$) for the Male Actor in Front and Side View Terms Are Sorted From High to Low Scores in Each View.

	Male								
	Front		Side			Front		Side	
1	Suspicious	0.989	Reflective	1	48	Defiant	0.829	Anxious	0.846
2	Intrigued	0.968	Baffled	0.975	49	Hostile	0.829	Cautious	0.845
3	Encouraging	0.961	Jealous	0.961	50	Regretful	0.826	Confused	0.845
4	Depressed	0.937	Puzzled	0.943	51	Relieved	0.826	Friendly	0.844
5	Despondent	0.934	Sarcastic	0.94	52	Curious	0.824	Decisive	0.843
6	Confident	0.934	Ashamed	0.925	53	Nervous	0.823	Concerned	0.842
7	Concerned	0.931	Stern	0.916	54	Reassuring	0.823	Comforting	0.84
8	Incredulous	0.926	Eager	0.915	55	Pensive	0.821	Earnest	0.838
9	Disappointed	0.906	Irritated	0.914	56	Hopeful	0.818	Arrogant	0.835
10	Sympathetic	0.905	Contemplative	0.913	57	Offended	0.815	Resentful	0.833
11	Convinced	0.902	Distrustful	0.909	58	Distrustful	0.813	Convinced	0.829
12	Indecisive	0.897	Suspicious	0.908	59	Indifferent	0.812	Uneasy	0.828
13	Dubious	0.896	Joking	0.904	60	Thoughtful	0.809	Deciding	0.827
14	Contented	0.893	Defiant	0.902	61	Playful	0.808	Perplexed	0.827
15	Eager	0.891	Confident	0.901	62	Dominant	0.799	Assertive	0.827
16	Friendly	0.891	Annoyed	0.9	63	Interested	0.796	Pensive	0.826
17	Cautious	0.89	Offended	0.897	64	Assertive	0.796	Embarrassed	0.825
18	Apologetic	0.888	Despondent	0.894	65	Perplexed	0.793	Accusing	0.824
19	Preoccupied	0.884	Intrigued	0.891	66	Doubtful	0.792	Insisting	0.822
20	Amused	0.883	Encouraging	0.891	67	Relaxed	0.791	Relaxed	0.82
21	Resentful	0.881	Affectionate	0.888	68	Insisting	0.784	Threatening	0.819
22	Jealous	0.881	Thoughtful	0.888	69	Guilty	0.769	Dominant	0.819
23	Sarcastic	0.88	Playful	0.885	70	Sceptical	0.762	Curious	0.814
24	Joking	0.88	Enthused	0.879	71	Fearful	0.761	Impatient	0.81
25	Alarmed	0.876	Preoccupied	0.879	72	Threatening	0.757	Imploring	0.809
26	Tentative	0.871	Worried	0.877	73	Flustered	0.749	Contented	0.809
27	Upset	0.871	Depressed	0.876	74	Desire	0.746	Indifferent	0.802
28	Earnest	0.868	Regretful	0.876	75	Fantasizing	0.744	Upset	0.801
29	Anticipating	0.866	Hostile	0.874	76	Dispirited	0.726	Insulting	0.801
30	Annoyed	0.864	Fascinated	0.874	77	Puzzled	0.723	Doubtful	0.794
31	Serious	0.858	Serious	0.873	78	Accusing	0.713	Guilty	0.793
32	Affectionate	0.857	Sympathetic	0.867	79	Arrogant	0.71	Apologetic	0.791
33	Deciding	0.854	Dispirited	0.866	80	Horrified	0.709	Bewildered	0.79
34	Decisive	0.853	Amused	0.866	81	Anxious	0.707	Fearful	0.785
35	Comforting	0.853	Entertained	0.862	82	Confused	0.705	Incredulous	0.779
36	Enthused	0.852	Anticipating	0.86	83	Impatient	0.705	Indecisive	0.77
37	Ashamed	0.851	Dubious	0.858	84	Bewildered	0.701	Alarmed	0.769
38	Entertained	0.844	Relieved	0.856	85	Uneasy	0.693	Hopeful	0.75
39	Baffled	0.84	Desire	0.853	86	Fascinated	0.693	Fantasizing	0.736
40	Stern	0.837	Grateful	0.853	87	Insulting	0.681	Flirtatious	0.732
41	Contemplative	0.836	Nervous	0.853	88	Worried	0.678	Aghast	0.685
42	Embarrassed	0.833	Interested	0.853	89	Satisfied	0.658	Reassuring	0.665
43	Imploring	0.831	Tentative	0.852	90	Aghast	0.617	Flustered	0.647
44	Flirtatious	0.831	Disappointed	0.852	91	Panicked	0.605	Horrified	0.614
45	Irritated	0.831	Sceptical	0.849	92	Hateful	0.6	Panicked	0.483
46	Grateful	0.83	Satisfied	0.848	93	Terrified	0.443	Terrified	0.471
47	Reflective	0.83	Hateful	0.848					

Table 7. Terms That Show a Difference Between the Observers' Decisions for the Female and Male Interpretations of Mental States and the Corresponding Results (% Correct) in Experiment 1.

	Male		Female	
Terms	Front	Side	Front	Side
Anticipating	30.3	48.48	36.36	78.79
Comforting	3.03*	12.12*	69.7	78.79
Confident	21.21*	21.21*	84.85	51.52
Contented	39.39	54.55	87.88	72.73
Decisive	45.45	42.42	18.18*	18.18*
Sarcastic	51.52	54.55	51.52	66.67

Note. Asterisks indicate performances that are statistically not better than chance (χ^2 – Yates correction for continuity; $\alpha > .05$).

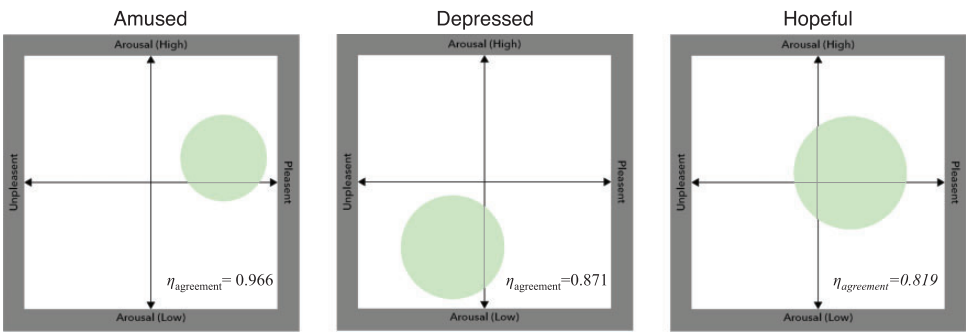

Figure 6. The figure shows a visualization of the magnitude of the agreement scores ($\eta_{agreement}$) within the valence-arousal space for three examples (amused, depressed, and hopeful). The radius of the green circles is equal to the values produced by Equation (1). The corresponding agreement scores ($\eta_{agreement}$) are represented in each graph.

correlations between the results for male faces in front view in Experiment 1 and male faces in front view in Experiment 2 ($r = -.302$, $p = .003$, $n = 93$), for male faces in front view in Experiment 1 and female faces in front view in Experiment 2 ($r = -.216$, $p = .038$, $n = 93$) and for male faces in side view in Experiment 1 and male faces in front view in Experiment 2 ($r = -.311$, $p = .002$, $n = 93$; see Table 8).

Discussion

Most currently available image databases of facial expressions of mental states include only a very small range of possible mental states. With the exception of the "Mind Reading" platform (Baron-Cohen et al., 2004), the vast majority of free databases employ the basic emotions proposed by Paul Ekman (1992; e.g., fear, disgust, surprise, happiness, sadness, and anger; see Table 1). Even the full set of emotions, however, constitute only one category of mental state to which ToM is directed. To investigate ToM comprehensively, a more expansive set of stimuli is desirable. The aim of the current study was to develop and to validate a new database of such stimuli reflecting a greater variety of mental states. The McGill Face Database includes four representations of 93 mental state terms. The pictures are unmodified but can be altered if users wish to do so. To determine the usefulness of the

database, two validation experiments were carried out. These experiments revealed considerable agreement among participants regarding the mental state expressed by the faces. Results from Experiment 1 demonstrate that subjects can reliably select the correct term associated with a particular mental state despite the sematic complexity of the terms denoting them. Subjects performed significantly better than chance in 78 of 93 front view images and 74 of 93 side view images of the female actor, and they performed significantly better than chance in 67 of 93 front view and 61 of 93 side view images of the male actor. Results from this experiment also show that subjects performed better with images of the female actor, most likely because she was more expressive than the male actor. It is noteworthy that while subjects' performance was better for front view images, the advantage over the side view was not dramatic (female: 84% vs. 80%; male: 72% vs. 66%). To our knowledge, this is the first demonstration of the high degree of accuracy human viewers exhibit when identifying complex mental states from only partially visible facial features. The Pearson correlation analyses for Experiment 1 show a highly significant correlation between the two views of the same face as well as between front views of the male and female faces. The slightly more difficult side view task together with differences across the male and female faces presumably accounts for the absence of the full complement of correlations. The aim of the validation in Experiment 2 was to develop a task that is independent of the complex vocabulary used in Experiment 1. This approach has a number of advantages. First, some of the mental state terms may be more likely to be chosen just in virtue of their meanings. These biases would distort subjects' performance. Second, the facial expressions produced by the actors are interpretations of mental state terms, and some interpretations may be more easily associated with a target term than others. In this respect, the relationship between the facial expressions and the mental state terms explored in Experiment 1 is distinctly different from the relationship between the basic emotions and the facial expressions to which they correspond. Although it is widely agreed that each basic emotion is represented by a single characteristic expression, many facial expressions might be thought to correspond to the mental state terms. Finally, it is of particular importance to be able to carry out ToM experiments without difficult vocabulary if one wants to study individuals with intellectual disabilities or those suffering from conditions associated with impaired linguistic ability. The "point-and-click" paradigm in which subjects had to indicate the location of a given facial expression in a logical space (Russell, 1980), along the dimensions of valence and arousal, makes this possible (Jennings et al., 2017). Results from this experiment show that there is substantial agreement across individuals about how to characterize faces along these dimensions. In addition, there is a high correlation between the face stimuli between perspectives and gender. The imperfect correlation between performance in the two experiments can be attributed to the presence of linguistic items in the first experiment and their absence in the second, as well as the difference in the specificity of the judgements required; the 2-D space used in Experiment 2 is a much coarser framework for classifying facial expressions than is the method of assigning a quite specific term to each face. The McGill Face Database thus provides a wide range of facial expressions of mental states that can be linked to mental state terms as well as accurately characterized in terms of arousal and valence independently of any such terms.

Acknowledgements

We would like to thank Prof. Erin Hurley from the Department of English at McGill University for her help with the audition and recruitment process of the actors and for her helpful comments in that process. We further thank Scott Brevity Cope (photographer) and Edward Schokking for their help.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a Social Sciences and Humanities Research Council of Canada grant #435–2017-1215, 2017 given to I. G. and G. S.

ORCID iD


Gunnar Schmidtmann https://orcid.org/0000-0001-8180-7751

Supplemental Material

Supplemental material for this article is available online.

References

Aifanti, N., Papachristou, C., & Delopoulos, A. (2010). The MUG facial expression database. In *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10* (pp. 1–4). IEEE.

Alyüz, N., Gökberk, B., Dibeklioğlu, H., Savran, A., Salah, A. A., Akarun, L., & Sankur, B. (2008, May). 3D face recognition benchmarks on the bosphorus database with focus on facial expressions. In *European workshop on biometrics and identity management* (pp. 57–66). Berlin, Heidelberg: Springer.

Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. (2004). *Mind reading: The interactive guide to emotions*. Retrieved from www.jkp.com

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*, 813–822.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The reading the mind in the eyes test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*, 241–251.

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence, 19*(7), 711–720.

Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: Meta-analysis. *Schizophrenia Research*, *109*, 1–9.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Brüne, M. (2005). Theory of mind in schizophrenia: A review of the literature. *Schizophrenia Bulletin*, *31*, 21–42.

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*, 351–362.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*, 169–200.

Georghiades, A. S., Belhumeur, P. N., & Kriegman, D. J. (2000). From few to many: Generative models for recognition under variable pose and illumination. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)* (pp. 277–284). IEEE.

Goeleven, E., Raedt, R. D., Leyman, L., & Verschuere, B. (2008). The Karolinska directed emotional faces: A validation study. *Cognition and Emotion*, *22*, 1094–1118.

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2010). Multi-pie. *Image and Vision Computing*, *28*, 807–813.

Harrington, L., Siegert, R., & McClure, J. (2005). Theory of mind in schizophrenia: A critical review. *Cognitive Neuropsychiatry*, *10*, 249–286.

Jennings, B. J., Yu, Y., & Kingdom, F. A. A. (2017). The role of spatial frequency in emotional face classification. *Attention, Perception, & Psychophysics*, *79*, 1573–1577.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3? *Perception*, *36*, 1–16.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. (2010). Presentation and validation of the Radboud face database. *Cognition & Emotion*, *24*, 1377–1388.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops* (pp. 94–101). Piscataway, NJ: IEEE.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 91*(630), 2–2.

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with gabor wavelets. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 200–205). Piscataway, NJ: IEEE.

Martinez, A. M. & Benavente, R. (1998). *The AR face database* (CVC Technical Report 24).

Min, R., Kose, N., & Dugelay, J.-L. (2014). Kinectfacedb: A kinect database for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*, 1534–1548.

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, & Computers*, *36*, 634–638.

Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005). Web-based database for facial expression analysis. In *2005 IEEE International Conference on Multimedia and Expo* (p. 5). Piscataway, NJ: IEEE.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161.

Schmidtmann, G., Sleiman, D., Pollack, J., & Gold, I. (2016). Reading the mind in the blink of an eye-a novel database for facial expressions. *Perception*, *45*, 238–239.

Setty, S., Husain, M., Beham, P., Gudavalli, J., Kandasamy, M., Vaddi, R., . . . Jawahar, C. V. (2013). Indian movie face database: A benchmark for face recognition under wide variations. In *2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)* (pp. 1–5). Piscataway, NJ: IEEE.

Sprong, M., Schothorst, P., Vos, E., Hox, J., & Van Engeland, H. (2007). Theory of mind in schizophrenia: Meta-analysis. *The British Journal of Psychiatry*, *191*, 5–13.

Tcherkassof, A., Dupré, D., Meillon, B., Mandran, N., Dubois, M., & Adam, J.-M. (2013). Dynemo: A video database of natural facial expressions of emotions. *The International Journal of Multimedia & Its Applications*, *5*, 61–80.

Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M., & Movellan, J. (2009). Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*, 2106–2111.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *Supplement to the Journal of the Royal Statistical Society*, *1*, 217–235.

Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006). A 3d facial expression database for facial behavior research. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)* (pp. 211–216). Piscataway, NJ: IEEE.

Appendix

Table A1. Parametric Pearson Correlations.

	Exp_1_ female_front	Exp_1_ female_side	Exp_1_ male_front	Exp_1_ male_side	Exp_2_ female_front	Exp_2_ female_side	Exp_2_ male_front	Exp_2_ male_side	
Exp_1_female_front	1	.555**	.336**	.201	–.087	–.09	–.123	–.131	
Sig. (two-tailed)		0	.001	.053	.408	.389	.239	.211	
N	93	93	93	93	93	93	93	93	
Exp_1_female_side	.555**	1	.157	.193	–.168	–.078	–.145	–.089	
Sig. (two-tailed)	0		.133	.064	.107	.46	.167	.397	
N	93	93	93	93	93	93	93	93	
Exp_1_male_front	.336**	.157	1	.598**	–.216*	–.145	–.302**	–.125	
Sig. (two-tailed)	.001	.133		0	.038	.167	.003	.232	
N	93	93	93	93	93	93	93	93	
Exp_1_male_side	.201	.193	.598**	1	–.175	–.163	–.311**	–.178	
Sig. (two-tailed)	.053	.064	0		.093	.119	.002	.089	
N	93	93	93	93	93	93	93	93	
Exp_2_female_front	–.087	–.168	–.216*	–.175	1	.520**	.436**	.321**	
Sig. (two-tailed)	.408	.107	.038	.093		0	0	.002	
N	93	93	93	93	93	93	93	93	
Exp_2_female_side		–.09	–.078	–.145	–.163	.520**	1	.391**	.297**
Sig. (two-tailed)	.389	.46	.167	.119	0		0	.004	
N	93	93	93	93	93	93	93	93	
Exp_2_male_front		–.123	–.145	–.302**	–.311**	.436**	.391**	1	.519**
Sig. (two-tailed)	.239	.167	.003	.002	0	0		0	
N	93	93	93	93	93	93	93	93	
Exp_2_male_side		–.131	–.089	–.125	–.178	.321**	.297**	.519**	1
Sig. (two-tailed)	.211	.397	.232	.089	.002	.004	0		
N	93	93	93	93	93	93	93	93	

Note. *Correlation is significant at the .05 level (two-tailed); **correlation is significant at the .01 level (two-tailed).