Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: http://www.tandfonline.com/loi/pcem20

Contributions of facial expressions and body language to the rapid perception of dynamic emotions

Laura Martinez, Virginia B. Falvello, Hillel Aviezer & Alexander Todorov

COGNITION AND EMOTION, 2016
Vol. 30, No. 5, 939–952, http://dx.doi.org/10.1080/02699931.2015.1035229

Routledge
Taylor & Francis Group

Contributions of facial expressions and body language to the rapid perception of dynamic emotions

Laura Martinez[1], Virginia B. Falvello[1], Hillel Aviezer[2], and Alexander Todorov[1]

[1]Department of Psychology, Princeton University, Princeton, NJ, USA
[2]Department of Psychology, Hebrew University, Jerusalem, Israel

(Received 24 June 2014; accepted 24 March 2015)

Correctly perceiving emotions in others is a crucial part of social interactions. We constructed a set of dynamic stimuli to determine the relative contributions of the face and body to the accurate perception of basic emotions. We also manipulated the length of these dynamic stimuli in order to explore how much information is needed to identify emotions. The findings suggest that even a short exposure time of 250 milliseconds provided enough information to correctly identify an emotion above the chance level. Furthermore, we found that recognition patterns from the face alone and the body alone differed as a function of emotion. These findings highlight the role of the body in emotion perception and suggest an advantage for angry bodies, which, in contrast to all other emotions, were comparable to the recognition rates from the face and may be advantageous for perceiving imminent threat from a distance.

Keywords: Emotion; Perception; Face perception; Body perception.

Expressing and recognising emotions in others are key components of human social interactions (Brosch, Pourtois, & Sander, 2010). It would be advantageous, for example, to recognise the expression of disgust in others who have eaten a poisonous plant or an expression of anger in someone who is planning to issue a threat. Although much research has been conducted on facial expressions of emotion, the stimuli used in many studies do not capture the richness and complexity of emotional expressions. In particular, the majority of prior research on emotion perception has focused solely on the face without taking into account the contributions of the affective body context (e.g., Ambady & Weisbuch, 2011; Calder & Young, 2005; Ekman, 1993; Smith, Cottrell, Gosselin, & Schyns, 2005). However, an exclusive focus on the face is unwarranted, given the fact that body language conveys important emotional information (Bänziger & Scherer, 2010; Dael, Goudbeek, & Scherer, 2013; Dael, Mortillaro, & Scherer, 2012; de Gelder, 2006). Indeed, body language may convey compelling emotion cues which influence, and at times, even overwrite facial emotion recognition (Aviezer, Hassin, Bentin, & Trope, 2008a; Aviezer et al., 2008b; Aviezer, Trope, & Todorov, 2012; de Gelder et al., 2006).

Correspondence should be addressed to: Alexander Todorov, Department of Psychology, Princeton University, Princeton, NJ, USA. E-mail: atodorov@princeton.edu

A particularly intriguing aspect of emotional communication relates to differences between emotion channels (face vs. body) in the recognisability of specific expressive cues. Prior work examining the recognition of static facial expressions across different distances has established that expressions of sadness are suited to closer-range communication and are better recognised up close than from a distance. In contrast, expressions of happiness function as good communication signals and are well recognised even from a distance (Smith & Schyns, 2009). Interestingly, in the latter study, facial expressions of anger were not well detected from a distance despite the obvious advantage of detecting threat from a distance. However, natural viewing conditions include a body which may better convey anger distally due to its larger size and broad dynamic patterns (Dael et al., 2012, 2013)

Recent work using dynamic expressions compared the recognisability of face and body cues across different emotions and increasing degrees of deteriorated viewing conditions (Visch, Goudbeek, & Mortillaro, 2014). Specifically, participants viewed emotional displays from intact videos, silhouettes with details, broad forms with no details or light point coordinates. In contrast to other emotions, anger body expressions were exceptionally robust and well recognised even when the body stimuli were impoverished and contained objectively less information. These findings are in accordance with the notion that emotion recognition, which is critical for detecting impending danger, must also operate under suboptimal conditions such as darkness and distance. While the aforementioned study examined deterioration of stimulus quality, an additional factor that may influence recognition is the duration of exposure to the stimulus. In particular, it would be of interest to examine if the robust recognition of dynamic anger bodies would also hold under brief exposures.

Prior studies have presented static faces with bodies for fixed brief durations (Meeren, van Heijnsbergen, & de Gelder, 2005) or unlimited durations (Aviezer et al., 2008b), yet it is of specific interest to examine how face–body accuracy develops over time beginning with very short dynamic exposures. While static images of emotions are recognised above chance levels after as little as 40 ms exposure (van de Riet & de Gelder, 2008), prior studies using affective dynamic stimuli have used relatively lengthy video clips lasting anywhere from 2 to 8 seconds (Atkinson, Dittrich, Gemmell, & Young, 2004; Kret, Pichon, Grèzes, & de Gelder, 2011).

In this study, we constructed a set of dynamic stimuli to test the relative contributions of faces and bodies at varying temporal durations to emotion perception. Following the considerations of Bänziger and Scherer (2007), we used footage from actors portraying emotions. We used a rather unconstrained procedure in order to increase natural portrayals of emotion in the body and face. We also included all basic emotions, including disgust that is sometimes omitted (de Gelder & Van den Stock, 2011) as it is difficult to express and recognise without additional contextual paraphernalia.

We chose to present dynamic stimuli because the expression and perception of emotions in real life is dynamic – faces and bodies move in three-dimensional space and people view the entire movements, not a single snapshot. Static stimuli do not capture the dynamic patterns that may also influence emotion recognition. For example, dynamic facial expressions are better recognised than static still images, an effect attributed mostly to the perception of change from the neutral starting position of the face (Ambadar, Schooler, & Cohn, 2005). Additionally, broad dynamic movements of the face and body (e.g., rotations, horizontal and vertical fluctuations and direction of the movement) may serve as important cues for emotion perception (Dael et al., 2012).

EXPERIMENT 1

After producing the affective stimuli, we presented the dynamic expressions in three formats: intact videos with faces and bodies, videos with masked bodies and videos with masked faces. This method allowed us to assess the relative recognition of emotions from faces and bodies in isolation, and compare this recognition to the recognition from the intact videos with faces and bodies. We also manipulated the length of the video clips to explore the amount of information needed to accurately identify emotions. The length ranged from 250 to 4000 milliseconds. While we refer in the current and following experiments to the amount of information in the stimulus, it should be noted that our dependent measure was in fact recognition rate which may correlate with, but is not equivalent to, the objective amount of information provided by these stimuli.

Methods

Stimuli creation

Actors. Six actors (three male and three female) between the ages of 18 and 22 with some form of acting experience were recruited from the Princeton University student population via email. Actors did not receive specific instructions about the emotions they would portray prior to filming so that their expressions would be more spontaneous. Actors were filmed one at a time, did not see the performances of any other actor and received no visual feedback of their portrayals.

Recording. Actors were filmed against a white background using a Canon Vixia HF S200 digital video camera at 30 frames per second. The camera was approximately eight feet away from the actors and positioned at a height of four feet. Actors were filmed from the waist up and were allowed to touch and interact with the back wall if desired, although this was not explicitly stated. A semicircle with a radius of three feet and a centre for the starting position of the actor was outlined on the ground to prevent the actors from moving outside of the camera's view. All actors wore a black long sleeve shirt, black gloves and a black swim cap.

Actors were told that they would be filmed expressing six different emotions: anger, disgust, fear, happiness, sadness and surprise. They were instructed to imagine a situation in which they would be experiencing the emotion they would be portraying. They were also instructed to be as natural as possible. In order to prevent obstruction of the face, actors were instructed not to touch their face or move their hands in front of or near their face. The order that the actors performed the six emotions was randomised for each actor. Each individual emotion was filmed over the course of 4 minutes, with the actor solely expressing one emotion during that period. Within the 4-minute segment, actors enacted the single emotion several times. For example, the actor reacted to a scenario they imagined, then "reset", moved back to the starting position and repeated. Over the course of each 4-minute filming session, actors were instructed to gradually increase the intensity of the emotion they were expressing in order to increase realistic portrayal. Actors took short breaks between the 4-minute filming periods of different emotions.

Editing. All clips were edited using Sony Vegas Pro 9.0. From the 4-minute period of filming a particular emotion, a 4-second long clip that was representative of the actor's portrayal of that emotion was selected. All of the 4-second clips were selected from the final minute of the 4-minute acting period, as this was the period when both facial expressions and body movements were at a maximum. Each 4-second clip began with the actor momentarily at rest before beginning to express the emotion. Each actor provided six 4-second long clips (one for each emotion). These 4-second clips formed the basis for creating the manipulated video clips used in the experiment.

First, the audio of the 4-second clip was muted. Second, using the masking tool, the clip was edited to show only the face or only the body for the entire duration of the clip (see Figure 1 and Supplemental Videos). Third, shorter time lengths were created from the originally selected clip. Specifically, to create different durations, clips were cut down to 2 from 4 seconds, to 1 from 2 seconds, to 500 milliseconds from 1 second and to 250 milliseconds from 500 milliseconds. When cutting down clips, the end of the longer clip was always discarded, preserving the beginning of the clip. Therefore, the clips used from one actor for a particular emotion were identical, only differing in

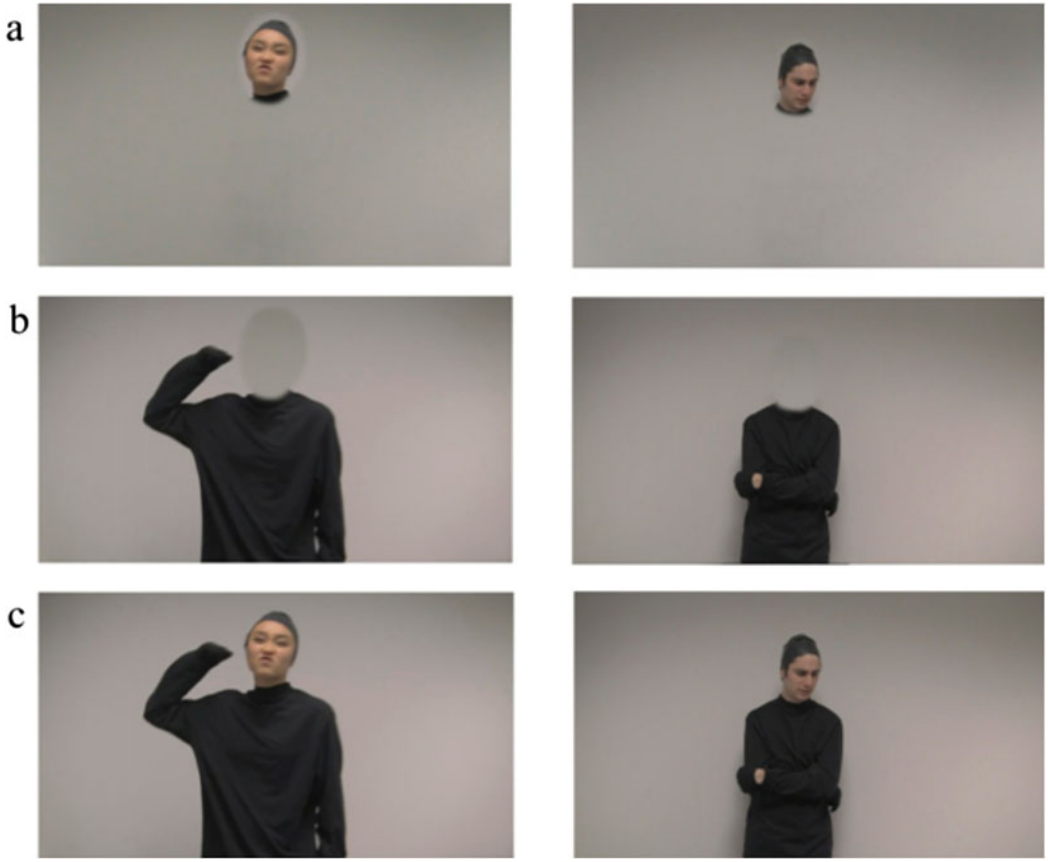

Figure 1. *Sample still frames from video clips of anger and sadness for the (a) face only, (b) body only and (c) face + body conditions.*

terms of their length (250, 500, 1000, 2000, 4000 milliseconds) and type (face, body, face + body). One second of black space was added to the beginning and end of each clip. A total of 540 clips were produced to use in the experiment (6 actors × 6 emotions × 3 types × 5 lengths).

Participants

Participants were recruited from the undergraduate student population at Princeton University. A total of 32 volunteers (20 female) participated in this study (18–22 years old, $M = 19.67$). Participants were compensated $12 for 1 hour of their time.

Procedure

The experiment was conducted using Media Lab version 2008.1.33 with a CRT monitor with refresh rate of 75 Hz. The clips were viewed from approximately 60 cm and subtended approximately 12° (vertically) and 4° (horizontally) for the intact face + body clips, 9° (vertically) and 4° (horizontally) for the body clips and 4° (vertically) and 1.5° (horizontally) for the face clips. Note that these values are only approximate and only apply to the fist snapshot in the clip, because the actors' movements change the visual angles. The 540 clips were grouped by length and type to form 15 blocks (5 lengths × 3 types, 36 videos in a block). Therefore, each block was defined by both the length of the clips and the type of stimuli it contained. The order of the blocks was randomised with the constraint that all "face + body" blocks (of intact videos) were always presented after the face blocks and the body blocks (of manipulated videos where the body or the face was masked, respectively) in order to prevent participants from recognising the face only and body only clips from their full version counterparts.

After viewing each clip, participants were shown a list of the six basic emotions (anger, disgust, fear, happiness, sadness and surprise) and were asked to choose which emotion the clip expressed. If unsure, participants were instructed to make their best guess. They were only allowed to view each clip once and were not allowed to choose more than one emotion in their responses. Participants could take brief breaks in between blocks as they desired.

Preliminary analyses

We computed both the standard (e.g., the proportion of identifying an emotion when present) and unbiased hit rates (Wagner, 1993; see also Hawk, van Kleef, Fischer, & van der Schalk, 2009). The latter controls for response biases where participants can predominantly use one or more emotion categories and are a better measure of recognition sensitivity. As outlined by Wagner (1993), we also computed the chance probability for each participant in each experimental condition. The unbiased hit rates were compared with these probabilities expected by chance. We not only report the analyses of the unbiased hit rates, but also note when these analyses are discrepant with analyses of the standard hit rates. Preliminary analyses showed that the findings are robust with respect to data transformations: the pattern of results and their significance was comparable for the raw proportions, their arcsine transformation, or their logit transformation (with imputing .01 to proportions of 0 and .99 to proportions of 1). Given that, we report the analyses of the raw proportions.

Table 1. *Raw (H) and unbiased hit rates (H_u) as a function of displayed emotion, emotion channel and exposure to dynamic emotion*

		Face		*Body*		*Face and body*	
Emotion	*Exposure (ms)*	*H*	H_u	*H*	H_u	*H*	H_u
Anger	250	.54	.46	.69	.37	.87	.78
	500	.63	.55	.78	.43	.88	.82
	1000	.60	.51	.79	.49	.91	.83
	2000	.71	.64	.92	.57	.93	.87
	4000	.71	.66	.92	.56	.94	.87
Disgust	250	.59	.44	.24	.15	.69	.61
	500	.65	.52	.25	.16	.71	.66
	1000	.71	.59	.34	.24	.73	.67
	2000	.80	.73	.30	.19	.84	.81
	4000	.85	.75	.33	.23	.86	.83
Fear	250	.66	.41	.39	.20	.65	.45
	500	.66	.47	.47	.28	.66	.58
	1000	.81	.56	.58	.35	.81	.63
	2000	.80	.59	.60	.38	.80	.69
	4000	.78	.58	.61	.40	.78	.75
Happiness	250	.95	.82	.27	.12	.97	.88
	500	.96	.85	.31	.16	.99	.91
	1000	.96	.86	.41	.25	.98	.91
	2000	.98	.94	.33	.22	.98	.91
	4000	.97	.94	.28	.19	.98	.96
Sadness	250	.85	.53	.61	.27	.85	.61
	500	.88	.61	.71	.31	.90	.67
	1000	.88	.66	.69	.32	.94	.75
	2000	.95	.74	.69	.31	.95	.80
	4000	.96	.76	.72	.32	.98	.81
Surprise	250	.61	.50	.34	.16	.67	.55
	500	.74	.58	.30	.17	.78	.67
	1000	.67	.56	.41	.26	.74	.66
	2000	.80	.73	.39	.28	.81	.77
	4000	.76	.70	.35	.23	.81	.77

Results

The unbiased hit rates were better than chance for all emotions at every time exposure and every clip type (face, body, face + body), *p*s < .001. Table 1 reports the standard and unbiased hit rates for all experimental conditions.

The unbiased hit rates were submitted to a 3 (clip type) × 6 (emotion) × 5 (exposure) repeated-measures analysis of variance (ANOVA).[1] This analysis found that all three main effects were highly significant, $F(2, 62) = 648.40$, $p < .001$,

[1] An additional analysis included gender of the participant as an additional between-subjects factor. This analysis did not find any effects related to gender.

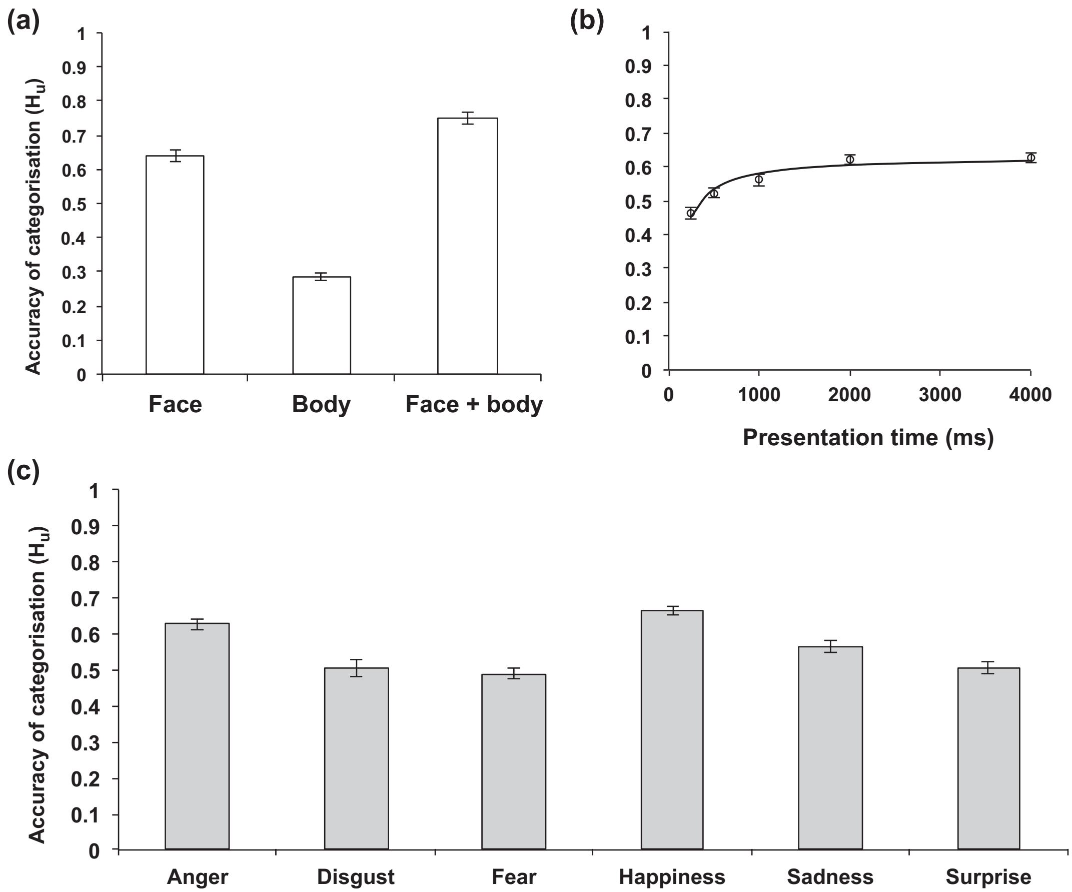

Figure 2. *Mean recognition accuracy (unbiased hit rates, H_u) of emotions as a function of (a) clip type (face, body, face + body), (b) exposure time and (c) type of emotion. Error bars represent standard error of the mean.*

$\eta^2 = .95$, for type of clip; $F(5, 155) = 48.08$, $p < .001$, $\eta^2 = .76$, for emotion; and $F(4, 124) = 95.27$, $p < .001$, $\eta^2 = .61$, for exposure time. As shown in Figure 2a, participants were least accurate for the body clips and, not surprisingly, most accurate for the intact (face + body) clips. As shown in Figure 2b, recognition accuracy rapidly increased with longer exposures and reached a plateau at 2 seconds' exposure. A sigmoid function accounted for 95% of the variance of the means for the different exposures (the function parameters, as well as the parameters for each emotion, are provided in Supplemental Table 1). Finally, participants were most accurate in recognising happiness and anger and least accurate in recognising fear, disgust and surprise (Figure 2c). Although the main effects were qualified by significant interactions, $F(8, 248) = 4.01$, $p < .001$, $\eta^2 = .12$, for clip type by exposure time; $F(10, 310) = 71.89$, $p < .001$, $\eta^2 = .70$, for clip type by emotion; $F(20, 620) = 3.15$, $p < .001$, $\eta^2 = .09$, for exposure time by emotion; and $F(40, 1240) = 2.85$, $p < .001$, $\eta^2 = .08$, for the three-way interaction), as shown in Table 1, the patterns of findings for the unbiased hits were similar across emotions. For every emotion, the lowest accuracy was achieved for the body cues and the highest for the intact videos and the accuracy increased with increases in exposure time to the emotion. Furthermore, all interactions except for the interaction of clip type and emotion were small in magnitude relative to the main effects.

We focus on the only large interaction of emotion and clip type, contrasting the unbiased hit rates for face and body cues as a function of emotion in a 6 (emotion) × 2 (face vs. body) repeated measures analysis. This analysis replicated the main effects for clip type, $F(1, 31) = 601.50$, $p < .001$, $\eta^2 = .95$, and emotion, $F(5, 155) = 27.55$, $p < .001$, $\eta^2 = .47$, and revealed an interaction effect, $F(5, 155) = 119.22$, $p < .001$, $\eta^2 = .79$. This interaction indicated that there were large differences between emotions in the recognition rates for the face and body cues, with the smallest difference for anger ($M = .07$, $SE = .02$) and the largest for happiness ($M = .69$, $SE = .02$). As shown in Table 1, anger was the only emotion where the findings for the raw hits and the unbiased hits were discrepant. For the raw hits measure, recognition was more accurate for body than for face cues, $t(31) = 8.43$, $p < .001$. However, for the unbiased hits, this difference reversed, $t(31) = 3.67$, $p < .001$. The reason for this discrepancy is that participants had a pronounced bias to report seeing anger when presented with bodies alone. They used this category 25% of the time. In contrast, they used the category 13% of the time when presented with faces alone and 15% when presented with intact videos.

Emotion confusability

In addition to the correct categorisation responses, we also analysed the emotion miscategorisation responses. An initial analysis of the data showed that the patterns of confusability were very similar across the different exposure times. The minimum (Pearson) correlation among the miscategorisation errors for the different exposure times was .72 for the face condition, .83 for the body condition and .80 for the face + body condition. The corresponding reliabilities in terms of Cronbach's alpha were .94, .98 and .96. Because of this similarity of the miscategorisation responses across different exposure times and for simplicity, we present the data aggregated across exposure times (Table 2).

Analysing the pattern of errors across emotions shows more similarity between the errors in the face and the face + body conditions, $r(30) = .82$, $p < .001$, than between the errors in the body and

Table 2. *Categorisation of emotions as a function of expressed emotion, perceived emotion and video clip*

		Perceived emotion					
Expressed emotion	*Video clip*	*Anger*	*Disgust*	*Fear*	*Happiness*	*Sadness*	*Surprise*
Anger	Face	**.64**	.09	.10	.02	.12	.04
	Body	**.82**	.02	.02	.06	.04	.05
	Face + body	**.91**	.01	.03	.01	.02	.01
Disgust	Face	.04	**.72**	.06	.00	.14	.05
	Body	.07	**.34**	.13	.06	.34	.07
	Face + body	.02	**.78**	.06	.00	.12	.03
Fear	Face	.03	.04	**.74**	.00	.12	.08
	Body	.04	.05	**.53**	.03	.26	.10
	Face + body	.01	.03	**.77**	.00	.12	.07
Happiness	Face	.00	.00	.00	**.97**	.02	.01
	Body	.26	.08	.03	**.32**	.19	.12
	Face + body	.00	.00	.00	**.98**	.01	.01
Sadness	Face	.01	.01	.01	.07	**.90**	.00
	Body	.05	.06	.13	.05	**.68**	.02
	Face + body	.00	.01	.01	.04	**.93**	.01
Surprise	Face	.04	.04	.17	.02	.01	**.72**
	Body	.25	.07	.10	.14	.08	**.36**
	Face + body	.06	.02	.12	.02	.01	**.76**

Note: Accurate categorisations (raw hits) are in bold font.

the face + body conditions, $r(30) = .56$, $p < .001$. The errors in the face and body conditions were weakly positively correlated, $r(30) = .23$, $p = .23$. In a regression analysis, both face errors (standardised $\beta = .73$, $p < .001$) and body errors (standardised $\beta = .40$, $p < .001$) predicted the face + body errors ($R^2 = .82$).

Although the errors in the face and body conditions were weakly correlated across all emotions, a more detailed analysis shows distinctive patterns within specific emotions. As shown in Table 2, the face and body errors were very similar for disgust and fear (correlations of .95 and .94, respectively). For example, disgust was most likely to be confused with sadness in both the face and body conditions. Fear too was most likely to be confused with sadness in these conditions. For the remaining emotions, the errors in the face and body conditions were quite different. For example, whereas a surprised face was most likely to be confused with a fearful face, a surprised body was most likely to be confused with an angry body. An angry face was most likely to be confused with a fearful or disgusted face, but an angry body was most likely to be confused with a happy body.

EXPERIMENT 2

Although Experiment 1 allowed us to study how recognition of emotions changes with increases in the amount of information from the onset of the expression, it also introduced a potential confound arising from the fact that samples from the same video clips were presented multiple times throughout the experiment. These multiple presentations could have facilitated the accurate identification of emotions. We sought to address this issue in Experiment 2, in which participants only viewed the shortest video clip. We focused on the 250 ms exposure, because the accurate identification of emotions in this condition was most surprising. Participants were randomly assigned to only one of three experimental conditions: face only, body only or face + body videos. Thus, each participant viewed each actor portraying each emotion only once and could not rely on exposure to other versions of the same video to inform their choices. We expected to observe the same pattern of responses as in Experiment 1 even with diminished overall identification of emotions.

Methods

Stimuli

From the videos created for Experiment 1, we used the thirty-six 250 ms clips of each actor portraying each emotion.

Participants

Sixty volunteers (41 female, ages 18–28 years old, $M = 20.42$) from the Princeton community participated for a payment of $8.

Procedure

The task was the same as in Experiment 1. Each clip was presented only once, and the viewing distance and visual angles were the same as in Experiment 1. Participants viewed the clips in random order in a single block. After viewing each clip, participants were asked to choose which one emotion the video expressed from a list of the six basic emotions. Participants were instructed to give their best guess if they were not sure which emotion was being portrayed.

Results

Replicating Experiment 1, the unbiased hit rates were better than chance for all emotions and every clip type (face, body, face + body), ps < .02. The unbiased hit rates were submitted to a 3 (clip type) × 6 (emotion) mixed-measures ANOVA with the first factor as between subjects and the second as within subjects[2] (see Table 3). As in Experiment 1,

[2] An additional analysis included gender of the participant as an additional between-subjects factor. This analysis found a small significant effect of gender, $F(1, 54) = 4.07$, p = .049, $\eta^2 = .07$, indicating that females ($M = .56$, $SE = .014$) were more accurate than males ($M = .51$, $SE = .022$). Gender did not interact with any of the other factors.

Table 3. *Raw (H) and unbiased hit rates (H_u) as a function of displayed emotion and emotion channel*

	Face		*Body*		*Face and body*	
Emotion	*H*	H_u	*H*	H_u	*H*	H_u
Anger	.38	.25	.42	.19	.58	.48
Disgust	.45	.24	.26	.11	.50	.36
Fear	.53	.31	.43	.19	.72	.43
Happiness	.97	.79	.25	.10	.97	.82
Sadness	.78	.47	.37	.15	.77	.49
Surprise	.59	.40	.32	.13	.60	.44

participants were most accurate at identifying emotions from the intact face + body clips and least accurate from the body clips (Figure 3a), $F(2, 57) = 78.40$, $p < .001$, $\eta^2 = .73$. The main effect of emotion was also significant, $F(5, 285) = 49.94$, $p < .001$, $\eta^2 = .47$, with highest accuracy for identification of happiness and lowest for disgust (Figure 3b). These main effects were qualified by a significant interaction, $F(10, 285) = 20.18$, $p < .001$, $\eta^2 = .42$. The interaction remained significant when contrasting the face and body cues as a function of emotion (and excluding the intact videos from the analysis), $F(5, 190) = 37.16$, $p < .001$, $\eta^2 = .49$.

As in Experiment 1, this interaction indicated that there were large differences between emotions in the recognition rates for the face and body cues (see Table 3), with the smallest difference for anger ($M = .06$) and the largest for happiness

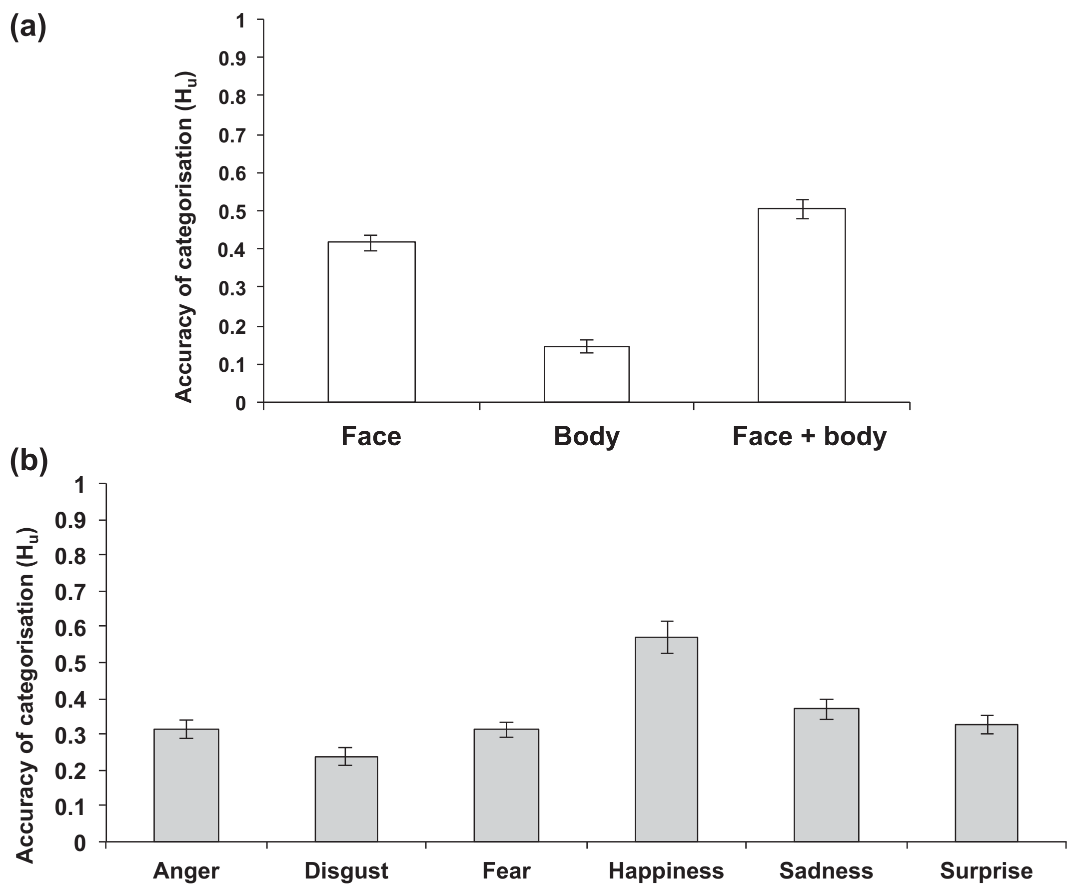

Figure 3. *Mean recognition accuracy (unbiased hit rates, H_u) of emotions by (a) clip type and (b) emotion. Error bars represent standard error of the mean.*

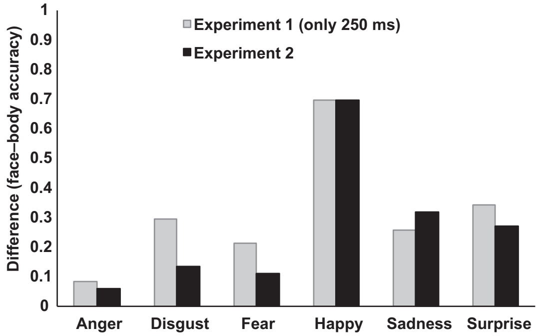

Figure 4. *Difference between the recognition rates (H_u) of emotions from face and body cues as a function of type of emotion in Experiments 1 and 2 (250 milliseconds exposure to emotion videos).*

(M = .70; note that it is not possible to compute standard errors for these differences, because the emotional cues were manipulated between subjects). As shown in Figure 4, despite the differences in experimental design, the pattern of differences was very similar across the two experiments. Again as in Experiment 1, anger was the only emotion where the findings for the raw hits and the unbiased hits were discrepant. Whereas for the raw hits recognition was more accurate for body than for face cues, for the unbiased hits recognition was more accurate for face than for body cues, although none of these differences reached significance. Replicating Experiment 1, participants had a bias to report seeing anger when presented with bodies alone, using this category 18% of the time (vs. 10% when presented with faces alone and 12% when presented with intact videos). Finally, the patterns of miscategorisation of emotions were also very similar across the two experiments, $r(90) = .80$, $p < 001$.

GENERAL DISCUSSION

Both the face and the body carry emotional information. While perceiving the face and body together resulted in the best recognition of basic emotions, both the face and the body alone provided enough information to identify an emotion above the chance level. Furthermore, very little information was needed to perceive dynamic emotions accurately. After as little as 250 ms exposure, participants were able to accurately identify emotions not only from the intact face + body clips but also from faces and bodies alone. Increasing the length of the videos improved recognition, but the benefits of increased length were diminishing. For all emotions, recognition accuracy plateaued at about 2 seconds.

There were also distinctive patterns for different emotions. For all emotions, the effects of exposure time on recognition depended on what type of video participants saw. Surprise, sadness, disgust and happiness followed a similar pattern: more time yielded more accurate recognition from the face + body and face conditions at a greater rate than from the body condition. For anger and fear, the two emotions where the advantage of the face cues over the body cues was the smallest (see Figure 4), more time yielded similar increases in the rate of accurate recognition for the face and body conditions. These findings suggest that the effect of time is more beneficial for information that is more diagnostic for identifying the respective emotion. When a channel is less diagnostic, additional time does not provide a great benefit for recognition.

Anger was the only emotion for which the face and body yielded comparable accuracy rates. In fact, for the uncorrected hits, angry bodies were more recognisable than angry faces. This body advantage in recognition of anger was eliminated when the analysis controlled for the use of different emotion categories. When participants saw bodies alone, but not faces alone or faces and bodies, they had a pronounced bias to see these bodies as angry. One possible explanation is that anger is different from the other emotions in that it represents a direct or imminent threat to the viewer, prompting an increased startle reflex and the "fight-or-flight" response (Springer, Rosas, McGetrick, & Bowers, 2007).

Indeed, fMRI studies indicate that perceiving whole-body expressions of anger elicit activity in regions including the amygdala and the lateral orbitofrontal cortex, which play a role in the affective evaluation of the stimuli. Further, the perception of dynamic anger bodies additionally engages brain regions that are coupled with autonomic reactions and motor responses related to defensive behaviours (Kret et al., 2011; Pichon, de Gelder, & Grezes, 2008). Because an expression of anger constitutes a direct threat, it would be advantageous to be able to recognise that emotion from a distance – rendering the body as a more useful source of information than the face. These results are also in good accordance with recent work showing that angry body postures are detected more rapidly than happy body postures in an analogue of the face-in-a-crowd effect (Gilbert, Martin, & Coulson, 2011). Our findings also complement the research by Visch et al. (2014) who found that recognition of anger from bodily expressions is highly robust with respect to stimulus degradation and more robust than recognition of fear, despair and joy.

Our current results demonstrate stark differences in recognisability between different channels of a given emotion. For example, happy faces were recognised near ceiling, with participants reaching near-perfect identification after only 500 ms exposure. However, recognition from the happy body was one of the lowest accuracies measured. The low recognition of the happy bodies echoes results from a study comparing static body emotions; happy bodies had relatively lower recognition rates than angry, fearful and sad bodies (de Gelder & Van den Stock, 2011). Despite these similar patterns, the overall recognition of emotion from our body stimuli was far lower than that found in prior studies and recently published sets (Coulson, 2004; de Gelder & Van den Stock, 2011; Thoma, Soria Bauser, & Suchan, 2013). Yet, rather than view this as a limitation of the current stimuli, we suggest that this reflects the holistic portrayal of emotion under natural emotional reenacting.

Nevertheless, several factors may have dampened the potential diagnosticity of the body in our study. First, we removed the heads from the body-only clips and excluding head motion may lead to an underestimation of the contribution of bodily expression. At the same time, head motion may have facilitated the correct identification of facial expressions. Second, our actors' body movements were limited by our request to avoid touching or moving the hands in the range of the face. While these constraints were included in order to allow comparison of the face and body, it is plausible that these factors further reduced the diagnostic information in the body condition. Some emotional expressions (e.g., sadness, disgust, fear) may inherently include the obstruction of the face by the hands (Atkinson et al., 2004; Atkinson, Tunstall, & Dittrich, 2007). Thus, our instruction may have hampered the natural reenactment of the body expressions. Future work retaining head motion (e.g., by blurring the face) and allowing for unconstrained emotional portrayals may reveal that the body is more diagnostic than our current data show. Lastly, our emotional portrayals were reenacted. Although the use of acted emotional expressions has been theoretically justified in previous work (Bänziger & Scherer, 2007, 2010; Scherer & Ellgring, 2007), such portrayals may nonetheless differ from spontaneous and naturally occurring expressions as imagining an emotional situation may entail different processes from actual emotional responding.

Notably, the categorisation errors for the face and body showed very different patterns. Isolated faces showed minimal confusability between positive and negative emotions. For example, happy faces were rarely misrecognised as conveying negative emotions. By contrast, happy bodies were misrecognised as conveying anger at nearly the same rate as they were correctly recognised, a finding which also emerged in recent studies (Kret, Stekelenburg, Roelofs, & de Gelder, 2013). Further, high body confusability was also found for surprise and anger. These patterns suggest that in highly active emotions the body conveys broad arousal/agitation information and little, if any, valence information. Only when the face and body are perceived together is the arousal of the body interpreted in accordance with the appropriate specific emotion. Interestingly, an unexpected confusability was found between disgust and sad bodies and between fear and sad bodies. These emotions share negative valence, but they convey quite different degrees of arousal and motor activation. One speculative possibility is that all these body expressions convey movements of withdrawal and avoidance. While disgust and fear may intuitively convey withdrawal, this may also be the case with sad body expressions (consider an individual staggering backwards upon hearing news causing grief and bereavement).

There is clearly redundancy in the information provided by the face and body channels, as indicated by the fact that recognition accuracy from the intact face + body videos was less than the sum of accuracy from the face only and body only conditions. Part of this redundancy could have emerged from the nature of the stimuli. Besides the information from the facial expressions in the face only videos, the position and broad movements of the head were visible, which also convey emotional information (Mignault & Chaudhuri, 2003). Furthermore, participants could infer if the face was moving towards or away from the camera based on the changes in head size, which provides information about the movement of the body. These considerations suggest that the findings in the face only condition were likely an overestimation of the information contained solely in facial expressions. However, the clips presented the face as it would appear in real life: with both dynamic facial expressions and head movement.

In typical face to face social interactions, explicit scrutiny of another's posture may not always be possible and may even be considered inappropriate. By contrast, when observing the expressions of a distant target, body language which is large and visually dominant may take precedence especially as some facial expressions are not easily decoded from a distance (Smith & Schyns, 2009). Because of these reasons, previous work comparing static faces and bodies presented the face proportionally larger than the body (Kret et al., 2013; van de Riet, Grèzes, & de Gelder, 2009). By contrast, in the current study, we presented the isolated bodies and faces in their original sizes, resulting in stimuli that differ significantly in size. While each methodological choice has its merits and limitations, an interesting consequence is that our finding that anger is as well recognised from the body as from the face may reflect the fact that anger is more difficult to decode from small, distant faces.

Finally, it is interesting to compare our results with those of a recent study comparing the recognition of affective faces and bodies (Aviezer et al., 2012). In that study, it was shown that during intense sport situations (specifically, victory vs. defeat), faces are typically non-diagnostic for the basic distinction of positive vs. negative valence while body language retained diagnosticity even during peak emotions (Aviezer et al., 2012). While the bodies in the current study retained diagnosticity (i.e., they were recognised above chance), faces were far more recognisable than in the aforementioned sports faces study. We suggest that the relatively high diagnosticity of the faces in the current study results from the moderate level of experienced emotions that the actors were able to reinstate and express in a controlled manner. By contrast, the levels of emotions in extreme real-life situations are likely to be far more intense and less controlled, leading to a breakdown in the diagnosticity of the face.

Supplemental data

Supplemental data for this article can be accessed here.

Disclosure statement

No potential conflict of interest was reported by the authors.

REFERENCES

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*, 403–410. doi:10.1111/j.0956-7976.2005.01548.x

Ambady, N., & Weisbuch, M. (2011). On perceiving facial expressions: The role of culture and context. In A. J. Calder, G. Rhodes, J. V. Haxby, & M. H. Johnson (Eds.), *Oxford handbook of face perception* (p. 479). Oxford: Oxford University Press.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*, 717–746. doi:10.1068/p5096

Atkinson, A. P., Tunstall, M. L., & Dittrich, W. H. (2007). Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition*, *104*(1), 59–72. doi:10.1016/j.cognition.2006.05.005

Aviezer, H., Hassin, R. R., Bentin, S., & Trope, Y. (2008a). Putting facial expressions into context. In N. Ambady & J. Skowronski (Eds.), *First impressions* (pp. 255–286). New York, NY: Guilford Press.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A.,… Bentin, S. (2008b). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*, 724–732. doi:10.1111/j.1467-9280.2008.02148.x

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*, 1225–1229. doi:10.1126/science.1224313

Bänziger, T., & Scherer, K. R. (2007). Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. In A. Paiva, R. Prada, & R. W. Picard (Eds.), *Affective computing and intelligent interaction, second international conference, ACII 2007, Lisbon, Portugal, September 12–14, 2007, Proceedings: LNCS, vol. 4738* (pp. 476–487). Berlin: Springer–Verlag.

Bänziger, T., & Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A source book* (pp. 271–294). Oxford: Oxford university Press.

Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorisation of emotional stimuli: A review. *Cognition & Emotion*, *24*, 377–400. doi:10.1080/02699930902975754

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*, 641–651. doi:10.1038/nrn1724

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, *28*(2), 117–139. doi:10.1023/B:JONB.0000023655.25550.be

Dael, N., Goudbeek, M., & Scherer, K. R. (2013). Perceived gesture dynamics in nonverbal expression of emotion. *Perception*, *42*, 642–657. doi:10.1068/p7364

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion*, *12*, 1085–1101. doi:10.1037/a0025737

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, *7*, 242–249. doi:10.1038/nrn1872

de Gelder, B., Meeren, H. K. M., Righart, R., Stock, J., van de Riet, W. A. C., & Tamietto, M. (2006). Beyond the face: Exploring rapid influences of context on face processing. *Progress in Brain Research*, *155*, 37–48. doi:10.1016/S0079-6123(06)55003-4

de Gelder, B., & Van den Stock, J. (2011). The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Frontiers in Psychology*, *2*, 181. doi:10.3389/fpsyg.2011.00181

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*, 384–392. doi:10.1037/0003-066X.48.4.384

Gilbert, T., Martin, R., & Coulson, M. (2011). Attentional biases using the body in the crowd task: Are angry body postures detected more rapidly? *Cognition & Emotion*, *25*, 700–708. doi:10.1080/02699931.2010.495881

Hawk, S. T., van Kleef, G. A., Fischer, A. H., & van der Schalk, J. (2009). "Worth a thousand words": Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, *9*, 293–305. doi:10.1037/a0015178

Kret, M.E., Pichon, S., Grèzes, J., & de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *NeuroImage*, *54*, 1755–1762. doi:10.1016/j.neuroimage.2010.08.012

Kret, M. E., Stekelenburg, J. J., Roelofs, K., & de Gelder, B. (2013). Perception of face and body expressions using electromyography, pupillometry and gaze measures. *Frontiers in Psychology*, *4*, 28. doi:10.3389/fpsyg.2013.00028

Meeren, H. K. M., van Heijnsbergen, C. C. R. J., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences*, *102*, 16518–16523. doi:10.1073/pnas.0507650102

Mignault, A., & Chaudhuri, A. (2003). The many faces of a neutral face: Head tilt and perception of dominance and emotion. *Journal of Nonverbal Behavior*, *27*(2), 111–132. doi:10.1023/A:1023914509763

Pichon, S., de Gelder, B., & Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Social Neuroscience*, *3*, 199–212. doi:10.1080/17470910701394368

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, *7*(1), 113–130. doi:10.1037/1528-3542.7.1.113

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*, 184–189. doi:10.1111/j.0956-7976.2005.00801.x

Smith, F. W., & Schyns, P. G. (2009). Smile through your fear and sadness: Transmitting and identifying facial expression signals over a range of viewing distances. *Psychological Science*, *20*, 1202–1208. doi:10.1111/j.1467-9280.2009.02427.x

Springer, U. S., Rosas, A., McGetrick, J., & Bowers, D. (2007). Differences in startle reactivity during the perception of angry and fearful faces. *Emotion*, *7*, 516–525. doi:10.1037/1528-3542.7.3.516

Thoma, P., Soria Bauser, D., & Suchan, B. (2013). BESST (Bochum Emotional Stimulus Set)—A pilot validation study of a stimulus set containing emotional bodies and faces from frontal and averted view. *Psychiatry Research*, *209*(1), 98–109. doi:10.1016/j.psychres.2012.11.012

van de Riet, W. A. C., & de Gelder, B. (2008). Watch the face and look at the body! *Netherlands Journal of Psychology*, *64*, 143–151. doi:10.1007/BF03076417

van de Riet, W. A. C., Grèzes, J., & de Gelder, B. (2009). Specific and common brain regions involved in the perception of faces and bodies and the representation of their emotional expressions. *Social Neuroscience*, *4*(2), 101–120. doi:10.1080/17470910701865367

Visch, V. T., Goudbeek, M. B., & Mortillaro, M. (2014). Robust anger: Recognition of deteriorated dynamic bodily emotion expressions. *Cognition & Emotion*, *28*, 936–946. doi:10.1080/02699931.2013.865595

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28. doi:10.1007/BF00987006