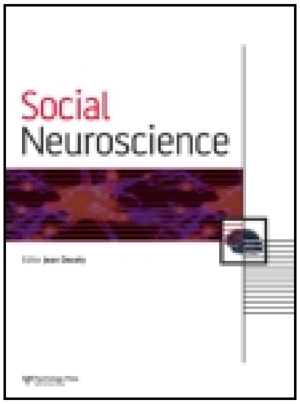

Social Neuroscience
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/psns20

Perceiving emotions from bodily expressions and multisensory integration of emotion cues in schizophrenia

Jan Van den Stock [a b] , Sjakko J. de Jong [a c] , Paul P. G. Hodiamont [a d] & Beatrice de Gelder [a b e]

[a] Cognitive and Affective Neuroscience Laboratory, Tilburg University, Tilburg, The Netherlands

[b] Department of Neuroscience, KU Leuven, Leuven, Belgium

[c] GGZ, Breburg Groep, Tilburg, The Netherlands

[d] Department of Developmental, Clinical and Cross-cultural Psychology, Tilburg University, Tilburg, The Netherlands

[e] Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Available online: 22 Jul 2011

SOCIAL NEUROSCIENCE, 2011, 6 (5–6), 537–547

Psychology Press
Taylor & Francis Group

Perceiving emotions from bodily expressions and multisensory integration of emotion cues in schizophrenia

Jan Van den Stock[1,2], Sjakko J. de Jong[1,3], Paul P. G. Hodiamont[1,4], and Beatrice de Gelder[1,2,5]

[1]Cognitive and Affective Neuroscience Laboratory, Tilburg University, Tilburg, The Netherlands
[2]Department of Neuroscience, KU Leuven, Leuven, Belgium
[3]GGZ, Breburg Groep, Tilburg, The Netherlands
[4]Department of Developmental, Clinical and Cross-cultural Psychology, Tilburg University, Tilburg, The Netherlands
[5]Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Most studies investigating emotion recognition in schizophrenia have focused on facial expressions and neglected bodily and vocal expressions. Furthermore, little is known about affective multisensory integration in schizophrenia. In the first experiment, the authors investigated recognition of static, face-blurred, whole-body expressions (instrumental, angry, fearful, and sad) with a two-alternative, forced-choice, simultaneous matching task in a sample of schizophrenia patients, nonschizophrenic psychotic patients, and matched controls. In the second experiment, dynamic, face-blurred, whole-body expressions (fearful and happy) were presented simultaneously with either congruent or incongruent human or animal vocalizations to schizophrenia patients and controls. Participants were instructed to categorize the emotion expressed by the body and to ignore the auditory information. The results of Experiment 1 show an emotion recognition impairment in the schizophrenia group and to a lesser extent in the nonschizophrenic psychosis group, and this for all four expressions. The findings of Experiment 2 show that schizophrenia patients are more influenced by the auditory information than controls, but only when the auditory information consists of human vocalizations. This shows that schizophrenia patients are impaired in recognizing whole-body expressions, and they show abnormal affective multisensory integration of bimodal stimuli originating from the same source.

Keywords: Schizophrenia; Body; Emotion; Audiovisual.

An important aspect of normal social functioning consists of recognizing intentions and emotions displayed by others. Emotion recognition in schizophrenia is hard to assess due to limited tools and studies that have predominantly focused on facial expressions (e.g., Borod, Martin, Alpert, Brozgold, & Welkowitz, 1993; Feinberg, Rifkin, Schaffer, & Walker, 1986; Heimberg, Gur, Erwin, Shtasel, & Gur, 1992; Kee, Horan, Wynn, Mintz, & Green, 2006; Kohler et al., 2003; Pomarol-Clotet et al., 2010; Wolwer, Streit, Polzer, & Gaebel,

Correspondence should be addressed to: Beatrice de Gelder, Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Room 417, Building 36, First Street, Charlestown, MA 02129, USA. E-mail: degelder@nmr.mgh.harvard.edu

We are grateful to all participants, especially the patients for their willingness to volunteer for the study. Research was funded by a Human Frontiers Science Program grant (RGP0054/2004-C), a European Commission grant (FP6-2005-NEST-Path Imp 043403-COBOL) to B.d.G., and a NWO grant (Dutch Science Foundation).

www.psypress.com/socialneuroscience http://dx.doi.org/10.1080/17470919.2011.568790

1996) (for a review, see Mandal, Pandey, & Prasad, 1998). The findings point to a deficit in recognition of negative emotions (Mandal et al., 1998), and this has been linked to the social dysfunctions observed in schizophrenia patients (Pinkham, Hopfinger, Ruparel, & Penn, 2008b). From that perspective, a facial expression recognition deficit is not very surprising, but an important issue is whether one can generalize from a deficit in recognition of facial expressions to difficulties in recognizing emotional signals conveyed by other common channels like the voice and the body. With the exception of a few isolated reports (Argyle, 1988; Sprengelmeyer et al., 1999), the literature on how body expressions are processed has only taken off in the last decade. One of the first basic research questions concerned whether observers can easily recognize different emotional states from body expressions alone. The available data indicate that this is clearly the case (Van den Stock, Righart, & de Gelder, 2007), and this is not surprising, considering the high frequency of interactions with conspecifics. Repeated exposure to body language, be it emotional or neutral, leads to perceptual expertise and tuning of the visual system. When we investigate emotional body language, comparing the results with what is known from facial expression research is almost inevitable, considering the many similarities between both stimulus categories. Bodies and faces both provide information on diverse dimensions such as identity, emotion, gender, and age (de Gelder et al., 2010). An interesting approach in comparing findings from face and body research might be to focus on the differences rather than on the similarities. At face value, at least two significant differences between faces and bodies appear.

Firstly, faces provide significantly more information about identity than bodies. Headless bodies reveal little information about personal identity, whereas faces alone are sufficient for identification. The fact that bodies contain little identity information is related to the fact that bodies are usually clothed. Clothing may conceal bodily features that are sufficient for identification. On the other hand, it has been shown that people can recognize friends by dynamic information provided by the body alone (Cutting & Kozlowski, 1977), underscoring the importance of dynamic information conveyed by body expressions. Secondly, the emphasis on the function of facial expressions lies in communication, whereas whole-body expressions also serve adaptive behavioral functions, like fight or flight.

Only a few studies have investigated the perception of whole-body expressions in clinical populations that have been shown to display facial expression recognition deficits, as in Huntington's disease, autism, and prosopagnosia (de Gelder, Van den Stock, de Diego Balaguer, & Bachoud-Levi, 2008; Hadjikhani et al., 2009; Tamietto, Geminiani, Genero, & de Gelder, 2007; Van den Stock, van de Riet, Righart, & de Gelder, 2008b). The results point to similar mechanisms for both categories. However, no data have been reported so far regarding whole-body expression perception in schizophrenia.

Another understudied area in affective neuroscience concerns how emotional information conveyed by different sensory channels is integrated (de Gelder & Van den Stock, 2011). The few studies so far have focused on the combined perception of face–voice pairs. The ability to decode emotional cues in prosody and facial expressions may have a common processing and/or representational substrate in the human brain (Borod et al., 2000; de Gelder & Bertelson, 2003; Pourtois, Debatisse, Despland, & de Gelder, 2002), facilitating processing and integration of these distinct but often calibrated sources of information. Judging the emotional state of a speaker is possible via facial or vocal cues (Banse & Scherer, 1996; Scherer, Banse, Wallbott, & Goldbeck, 1991) alone, but both judgment accuracy and speed seem to benefit from combining the modalities; for example, response accuracy increases and response speed decreases when a face is paired with a voice expressing the same emotion. This improvement of performance occurs even when participants are instructed to ignore the voice and rate only the face, suggesting that extracting affective information from a voice may be automatic and/or mandatory (de Gelder & Vroomen, 2000). The fact that prosodic and facial expressions of emotion frequently correlate suggests that the underlying cognitive mechanisms are highly sensitive to shared associations activated by cues in each channel (de Gelder, Bocker, Tuomainen, Hensen, & Vroomen, 1999; Massaro & Egan, 1996).

To assess how emotional judgments of the face are biased by prosody, Massaro and Egan (1996) presented computer-generated faces expressing a *happy, angry,* or *neutral* emotion accompanied by a word spoken in one of the three emotional tones. De Gelder and Vroomen (2000) presented photographs taken from the Ekman and Friesen (1976) series with facial expressions rendered emotionally ambiguous by "morphing" the expressions between *happy* and *sad* as the two endpoints. The emotional prosody tended to facilitate how accurately and quickly subjects rate an emotionally *congruent as compared to an incongruent* face. These findings indicate that the emotional value of prosody–face events is registered and somehow integrated during perceptual tasks, affecting behavioral responses according to the emotion congruity of the combined events. Moreover, these cross-modal influences appear to be resistant to increased attentional demands induced by a dual task, implying that combining the two forms of input may be mandatory (Vroomen, Driver, & de Gelder, 2001). Recent studies have only begun to unravel the mechanisms behind cross-modal influences on perception of whole-body expressions, and the results are compatible with what has been previously reported for facial expressions (Van den Stock, Grèzes, & de Gelder, 2008a; Van den Stock, Peretz, Grèzes, & de Gelder, 2009; Van den Stock et al., 2007). However, no evidence has been reported so far on cross-modal influences on whole-body perception in schizophrenia. We recently used affective face–voice combinations that were either congruent (for example, a happy face presented simultaneously with a happy vocal expression) or incongruent (for example, a happy face paired with a fearful vocal expression), and we asked schizophrenia patients to rate one of the two modalities (de Gelder et al., 2005; de Jong, Hodiamont, Van den Stock, & de Gelder, 2009). The results showed anomalous, cross-modal bias effects in the patient group. For example, when schizophrenia patients were instructed to categorize the emotion expressed in the voice, they were less influenced than the controls by the facial expression (de Jong et al., 2009).

In the present study, we focus on the perception of whole-body expressions. In Experiment 1, we investigated the recognition of emotional body language in a group of schizophrenia patients, nonschizophrenic psychotic patients, and normal controls in order to determine whether the emotion-recognition deficit previously reported for faces (for a review, see Mandal et al., 1998) extends to whole-body expressions. So far, little is known about recognition of emotional body language in schizophrenia, but in view of the behavioral and neuroanatomical similarities between perception of faces and bodies in normals (for reviews, see de Gelder, 2006; de Gelder et al., 2010; Peelen & Downing, 2007), we hypothesized that the patients would be impaired in recognizing negative whole-body expressions.

In Experiment 2, we investigated how schizophrenia perceive multisensory emotional events, consisting of realistic body language combined with affective auditory utterances. We presented video clips of emotional body language, of people engaged in a common activity in an everyday situation. In addition to adding human vocal expressions, we also combined the video clips with animal vocalizations in order to investigate the role of environmental sounds. As reported previously, it is important to control for task variables, as attention may shift across conditions and trials from face to voice especially in clinical populations (Bertelson & de Gelder, 2004; de Gelder & Bertelson, 2003).

EXPERIMENT 1: RECOGNITION OF STATIC BODY LANGUAGE

Methods and materials

Participants

Thirty-one schizophrenia, 23 patients with nonschizophrenic psychosis, and a group of 21 normal controls matched for gender, age, and socioeconomic status participated in the study. The nonschizophrenic psychosis group consisted of patients with schizophreniform disorder ($n = 2$), schizoaffective disorder ($n = 5$), bipolar I disorder with psychosis ($n = 4$), depressive disorder with psychosis ($n = 1$), delusional disorder ($n = 1$), psychotic disorder not otherwise specified ($n = 9$), and dysthymic disorder ($n = 1$). All but three patients in the nonschizophrenic psychosis group and all but three patients in the schizophrenia group received antipsychotic medication. Demographic data are shown in Table 1. There was no significant difference in age, $F(2, 72) = 0.620$; $p < .541$, or gender ratio, $\chi^2 \leq 2.13$; $p < .14$, between the three groups. Only patients meeting the criteria

TABLE 1
Demographic data

Demographic data	*Schizophrenia*	*Nonschizophrenic psychosis*	*Normal*
Experiment 1			
N	31	23	21
Age (mean range)	33.7 (21-52)	35.7 (20-54)	32.4 (21-58)
Gender	23 M/8 F	14 M/9 F	13 M/8 F
Dexterity	28 R/3 L	20 R/3 L	20 R/1 L
Experiment 2			
N	16	/	16
Age (mean range)	36.8 (22-53)	/	38.0 (22-53)
Gender	15 M/1 F	/	9 M/7 F
Dexterity	15 R/1 L	/	13 R/3 L

for schizophrenia and nonschizophrenic psychosis set by the DSM-IV (APA, 2000) were included. All patients were under treatment at the local hospital. Diagnosis was established with the Schedules for Clinical Assessment in Neuropsychiatry (SCAN, version 2.1), a standardized interview for diagnosing axis I disorders, conducted by a trained psychiatrist. Exclusion criteria for patients consisted of: organic or substance-induced psychosis, current substance abuse, serious somatic illness, relevant neurological illness, auditory and /or visual handicap and language problems. Control subjects with a psychiatric disorder, a brain dysfunction, or a genetic predisposition to schizophrenia were excluded from participation. All participants were paid for participation (€22).

Materials and procedure

Materials consisted of pictures from our own database of body expressions and instrumental actions (for details on stimulus construction, see de Gelder, Snyder, Greve, Gerard, & Hadjikhani, 2004; Hadjikhani & de Gelder, 2003). The faces of the bodies were blurred. In a pilot study, the pictures were presented one by one for 4 s with an interstimulus interval of 7 s. Twenty participants were instructed to categorize the pictures according to expressed emotion by indicating one of three (anger, fear, sadness) response alternatives. For the instrumental actions, the participants were instructed to categorize the displayed action (combing hair, drinking, pouring water in glass, opening door, talking on telephone, or putting on trousers). Only pictures that were correctly recognized above 85% were selected for the experiment.

The experiment consisted of two blocks: one with bodily expressions and one with bodily actions. We included instrumental whole-body actions, because these displays elicit action representation (Johnson-Frey et al., 2003) and are thus appropriate to use as controls to investigate emotional body expressions. The procedure was identical in each block. Materials for the experiment consisted of 30 emotional bodies and 24 instrumental actions. A stimulus consisted of the presentation of a target at the top of the screen that had to be matched with one of two simultaneously presented probes underneath (Figure 1). The three pictures in a stimulus were always three different identities, but from the same gender. The instructions stated that the participant was to select the probe that matched the action or emotion of the target. The position of the correct probe was counterbalanced. Participants responded by pressing the corresponding button, indicating their choice for the left or right probe. The stimulus was presented until the participant responded. During the 1000-ms intertrial interval, a blank screen was shown. The instrumental action block consisted of 48 trials (6 actions × 2 genders × 4 exemplars), and the bodily expression block consisted of 36 trials (3 expressions × 2 genders × 6 exemplars). For every emotion, 12 trials were presented, half with male images, half with female images. For example, the six male anger trials consisted of three trials where sadness was a distracter and three trials where fear was a distracter.

Results

Results are displayed in Figure 1. We calculated for every condition and participant the mean accuracy and median reaction times (RT) of the correct trials. Both RT and accuracy data were submitted to a repeated-measures ANOVA with Expression (four levels: angry, fearful, sad, and instrumental) as within-subjects factor and Group (three levels: schizophrenia, nonschizophrenic psychosis, and control) as between-subjects factor. This revealed for the accuracy data a significant main effect of Expression, $F(3, 210) = 13.269$; $MSE = 0.022$, $p < .001$, and Group, $F(2, 70) = 6.234$; $MSE = 0.033$, $p < .003$. The Expression × Group interaction was not significant. Tukey post hoc tests on the main effect of Group showed a significant difference between the control group and the schizophrenia group (mean difference = 0.093, $p < .002$), and a marginally significant difference between the control group and the nonschizophrenic psychosis group (mean difference = 0.067, $p < .054$). To follow up the main effect of Expression, we performed Bonferroni-corrected paired-sample *t*-tests between every combination of expressions ($n = 6$). This showed significant differences between angry and fearful, $t(74) = 5.911$; $p < .001$; between instrumental and fearful, $t(74) = 6.818$; $p < .001$; and between instrumental and sad expressions, $t(74) = 4.303$; $p < .001$. The difference between fearful and sad expressions was marginally significant, $t(74) = 2.483$; $p < .015$.

The analysis of RT showed a significant main effect of Expression, $F(3, 210) = 8.762$; $MSE = 116.649$, $p < .003$. The main effect of Group and the Expression × Group interaction were not significant. Bonferroni-corrected paired-sample post hoc *t*-tests showed a significant difference between angry and instrumental, $t(74) = 3.481$; $p < .001$; between fearful and sad, $t(74) = 3.326$; $p < .001$; between fearful and instrumental, $t(74) = 5.142$; $p < .001$; and between sad and instrumental, $t(74) = 4.009$; $p < .001$, expressions.

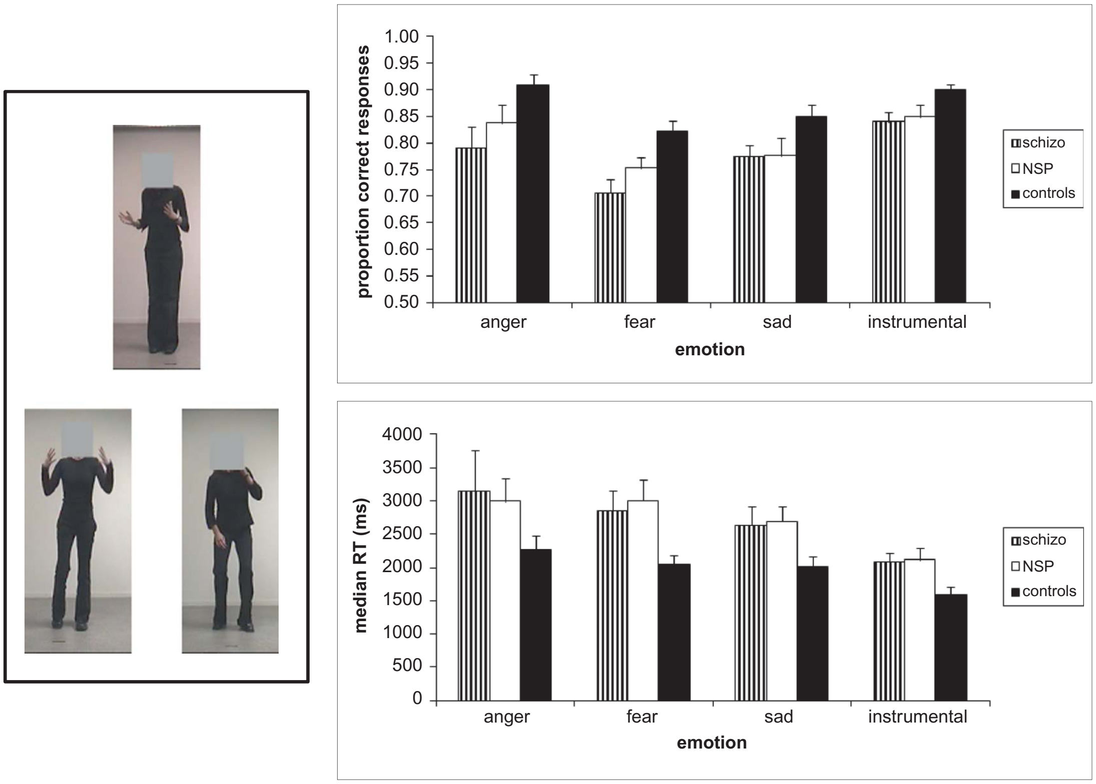

Figure 1. Left: example of stimulus in emotion block showing an angry body on top and bottom left and a fearful body on the bottom right. Right: Accuracy (top) and reaction time (bottom) of Experiment 1 as a function of expression and group (Schizo: schizophrenia group; NSP: nonschizophrenic psychosis group). Error bars represent 1 *SEM*.

Discussion

We presented patients with schizophrenia, patients with nonschizophrenic psychosis, and matched controls with a two-alternative, forced-choice, whole-body expression-matching task. The results show that, compared to the control group, the schizophrenia group exhibits a general impairment in recognizing emotional body language. The nonschizophrenic psychosis group occupies an intermediate position between the controls and schizophrenia group. It is unlikely that the observed effects can be explained by task difficulty, since the absence of an Expression × Group interaction reveals that the patients are not differentially impaired on recognition of specific emotions, while the main emotion effect indicates that not all emotions are equally well recognizable. Hence, the patients are not more impaired in recognizing the more difficult emotions.

The generalized whole-body emotion recognition deficit is consistent with findings from facial expression recognition studies (reviewed in Mandal et al., 1998). Schizophrenia patients are in general less able to make adequate emotional judgments of ambiguous facial expressions (Kee et al., 2006), and they attribute negative emotional valence to neutral face cues (Kohler et al., 2003). This may explain the nonspecific nature of emotion-recognition difficulties.

An important ability to adequately recognize facial expressions concerns the processing of the configuration of the face. It has been shown that this configurational perception mechanism is impaired in schizophrenia patients (Joshua & Rossell, 2009). Recent studies have reported similar configurational processing mechanisms for faces and bodies (Stekelenburg & de Gelder, 2004). The impaired recognition of whole-body expressions in schizophrenia may therefore have its roots in deficient configurational processing, and this may lead in social situations to inadequate interpretation of both facial and bodily expressions and ensuing social dysfunction.

EXPERIMENT 2: MULTISENSORY INTEGRATION OF DYNAMIC BODY LANGUAGE AND HUMAN AND ANIMAL VOCALIZATIONS

In everyday situations, fearful body language is usually accompanied by anxious screams. Recently, we showed that static whole-body expressions influence recognition of simultaneously presented vocal expressions (Van den Stock et al., 2007, experiment 3). In a follow-up study (Van den Stock et al., 2008a), we used dynamic stimuli in realistic situations to increase ecological validity, which may be an important factor in multisensory integration (Bertelson & de Gelder, 2004; de Gelder & Bertelson, 2003). We paired these visual stimuli with nonverbal vocalizations, and, more importantly, we also manipulated the nature of the bimodal combinations. We used auditory stimuli that were either produced by the same source as the visual stimuli (human vocalizations), or by a different source (animal vocalizations). The findings showed that both human and animal sounds influence recognition of dynamic body language. The second objective of the present study is to investigate the multisensory integration pattern of these everyday emotional events in schizophrenia.

Methods and materials

Participants

Sixteen schizophrenia meeting the criteria described in Experiment 1 and 16 matched controls participated in Experiment 2. All patients received antipsychotic medication. Demographic data are shown in Table 1. There was no significant difference of age between groups, $t(30) = 0.32$, $p < .751$. None of the participants of Experiment 2 participated in Experiment 1.

Materials and procedure

Visual stimuli. Video recordings were made of 12 semi-professional actors (6 women), coached by a professional director. They were instructed to approach a table, pick up a glass, drink from it, and put it back on the table. They performed this action once in a happy and once in a fearful manner. During the recording of the video clips, actors were provided with specific scenarios for every emotion; for example, the fearful scenario stated that the glass contained extremely hot water (see also Grèzes, Pichon, & de Gelder, 2007). A fragment of 800 ms showing the actor grasping the glass was selected from each take. Facial expressions were blurred by motion-tracking software. In a pilot study, the 24 edited dynamic stimuli were presented 4 times to 14 participants. Participants were instructed to categorize as accurately and as fast as possible the emotion expressed by the actor (fear or happiness). The pilot session was preceded by eight familiarization trials. Sixteen stimuli were selected (2 genders × 4 actors × 2 emotions). Since we expected that recognition of the body language would improve when the body stimuli were combined with congruent auditory information, body stimuli that were recognized at ceiling were not selected. Mean recognition of the selected stimuli was 86.1% ($SD = 9.7$). A paired t-test between the fearful and happy body language showed no significant difference, $t(13) = 1.109$, $p < .287$.

Auditory stimuli. Audio recordings were made at a sampling rate of 44.1 kHz of 22 subjects (14 women), while they made nonverbal emotional vocalizations (fearful and happy). Specific scripts were provided for every target emotion. For example, for fear, the subjects were instructed to imagine that they were going to be attacked by a robber. Explicit instructions were given to refrain from pronouncing words. The most representative 800 ms from each recording was selected. In a pilot study, the sounds were presented 4 times to 15 participants in a randomized order. The participants were instructed to categorize as accurately and as fast as possible the emotion expressed by the voice (fear or happiness). The pilot session was preceded by three familiarization trials. From these results, eight fearful and eight happy sounds were selected. Mean recognition of the stimuli was 94.6% ($SD = 6.7$). A paired t-test between the fearful ($M = 96.1$) and happy ($M = 93.0$) vocalizations showed no significant difference, $t(14) = 0.474$, $p < .643$.

Environmental sounds of aggressive dog barking and joyful bird songs were downloaded from the Internet. Stimuli were selected on the basis of their emotion-inducing characteristics. The most representative 800-ms fragment of every sound was presented 4 times to 13 participants in a third pilot study. Instructions were to categorize as accurately and as fast as possible the emotion induced by the sound (fear or happiness). The session was preceded by three familiarization trials. Eight fear- and eight happiness-inducing sounds were selected. Mean recognition of the stimuli was 94.8% ($SD = 5.7$). A paired t-test between the fearful ($M = 97.5$) and happy ($M = 92.1$) vocalizations showed no significant difference, $t(12) = 1.469$, $p < .168$.

For each emotion, we compared the ratings of the animal vocalizations with those of the human vocalizations. Independent samples *t*-tests showed no differences between the pairs, $t(26) \leq 1.195$, $p < .243$. Experimental stimuli were then constructed with these visual and auditory materials. For this purpose, every video was paired once with a fearful and a happy human vocalization and once with a fearful and a happy animal vocalization, resulting in a total of 64 bimodal stimuli.

Procedure. The experiment consisted of an auditory (A), visual (V), and audiovisual (AV) block. In each block, all stimuli were presented twice in random order. The order of the blocks was counterbalanced. The AV block consisted of 128 trials, the V block of 32 trials, and the A block of 64 trials. A trial started with the presentation of a white fixation cross in the center of the screen against a dark background. The fixation cross had a variable duration to reduce temporal predictability (2000–3000 ms) and was followed by presentation of a stimulus (800 ms), after which a question mark appeared until the participant responded. In the AV and V blocks, participants were instructed to categorize the emotion expressed by the body in a two-alternative, forced-choice task by pressing the corresponding button (happy or fearful). Response buttons were counterbalanced across participants. Because we wanted to make sure participants saw the full length of the stimulus before they responded, they were instructed to respond only when the question mark appeared. In the A block, participants were presented with only auditory stimuli and instructed to categorize the emotion (happy or fearful).

Results

We excluded trials on which participants responded before the end of the stimulus (RT < 800 ms). On this basis, 64 trials (1.3%) were discarded. We computed the proportion of "happy" responses in the different conditions. Results are shown in Figure 2. The data with animal and human vocalizations were analyzed separately. Since the participants performed a delayed-RT task, no RT data were analyzed. A comparison between both groups on recognition of each of the four unimodal auditory conditions showed no significant difference, $t(31) < 1.850$, $p < .074$.

Human vocalizations

A repeated-measures ANOVA was performed on the proportion of "happy" responses, with Visual Emotion (two levels: fearful and happy) and (human) Auditory Emotion (three levels: fearful, happy, and no auditory stimulus) as within-subjects factors, and with Group (two levels: schizophrenia and control) as between-subjects factor. This revealed a main effect of Visual Emotion, $F(1, 31) = 124.154$, $MSE = 0.102$, $p < .001$; a main effect of Auditory Emotion, $F(1, 31) = 11.278$, $MSE = 0.035$, $p < .001$; and a significant two-way Auditory Emotion × Group interaction, $F(2, 62) = 3.310$, $MSE = 0.035$, $p < .043$. The Visual Emotion × Group interaction was marginally significant, $F(1, 31) = 3.937$, $MSE = 0.102$, $p < .056$.

The main effect of Visual Emotion indicates that the proportion of happy responses is higher for happy body language, as expected. The main effect of Auditory Emotion shows that the ratings of the bodily expressions are influenced by Auditory Emotion, while the Auditory Emotion × Group interaction indicates that this auditory influence is significantly different between the two groups. To follow up this interaction effect, we computed the influence of the auditory information for both groups separately, by calculating the ordinal difference between the unimodal and bimodal conditions: (fear video minus fear video paired with fear audio) + (fear video paired with happy audio minus fear video) + (happy video paired with happy audio minus happy video) + (happy video minus happy video paired with fearful audio). The resulting difference was higher for the schizophrenia group (0.46) than the control group (0.14), indicating that the schizophrenia are more influenced by the vocalizations than the controls.

We also compared the ratings of both unimodal conditions (fearful and happy body language) between both groups with independent samples *t*-tests. This showed no significant difference, $t(31) < 1.335$, $p < .192$, indicating that both patients and controls were equally able to recognize the unimodal whole-body expression videos.

Animal vocalizations

A repeated-measures ANOVA on the proportion of happy responses with Visual Emotion (fearful and happy) and (animal) Auditory Emotion (fearful, happy, and no auditory stimulus) as within-subjects factors, and Group (schizophrenia and control) as between-subjects factor, revealed a significant main effect of Visual Emotion, $F(1, 31) = 112.758$, $MSE = 0.115$, $p < .001$, and a significant Visual Emotion × Group interaction, $F(1, 31) = 4.456$, $MSE = 0.115$, $p < .043$. To follow up the interaction effect, we computed for both groups separately, the mean proportion of happy responses for the conditions with a happy

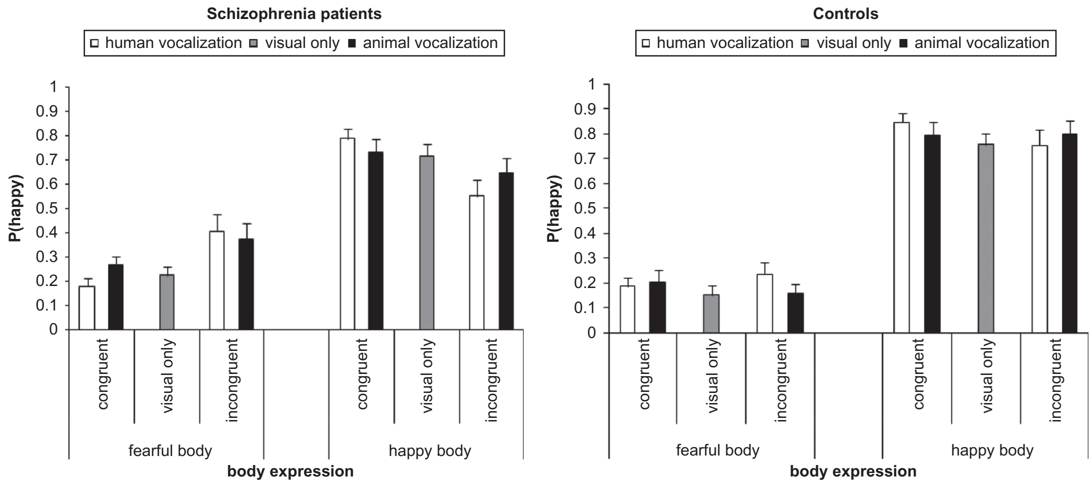

Figure 2. Proportion of “happy” responses in the bimodal and unimodal conditions, separated by group, emotion, auditory category, and congruence. Error bars represent 1 *SEM*.

video, regardless of (animal) auditory information, and we followed the same procedure for the fearful video conditions. The proportion of happy responses on the conditions with fear videos was significantly lower in the control group than in the schizophrenia group, $t(31) = 2.199$, $p < .035$), suggesting that the schizophrenia patients were less accurate in categorizing fearful videos. The difference between both groups in the conditions with happy videos was not significant.

Discussion

We presented schizophrenia patients and controls with videos of a person engaged in an everyday action (picking up a glass in a realistic situation). This action was performed in either a fearful or a happy manner. In the bimodal blocks, the videos were simultaneously presented with either a congruent or an incongruent vocal expression, which could be produced by a human or an animal. These stimuli were chosen to maximize ecological validity, which is an important factor in multisensory integration. It has been suggested that integrating information from multiple sensory channels leads to reduction of the ambiguity that is inherent in each single sensory channel (Bertelson & de Gelder, 2004; de Gelder & Bertelson, 2003). The results show that schizophrenia patients are more influenced by the task-irrelevant auditory information than the control group, but only for human and not for the animal vocalizations. The increased cross-modal bias of vocal expressions over body expressions may point to a greater impact of the auditory modality under audiovisual perception conditions in schizophrenia patients. This explanation is also compatible with our previous study in which schizophrenia patients showed a reduced cross-modal bias of visual facial expression in the recognition of the emotion in a vocal expression (de Jong et al., 2009), and with a recent report from audiovisual speech perception in schizophrenia (Ross et al., 2007). But this does not explain why schizophrenia patients are more influenced by the human and not the animal vocalizations. The present results indicate that abnormal affective multisensory integration in schizophrenia is modulated by the association between the sources of different sensory channels.

Another possible explanation may be task difficulty. If schizophrenia patients have more difficulty in recognizing the visual stimulus, they might rely more on the information provided by the secondary stimulus. A direct test between both groups of the unimodal conditions reveals no significant difference, indicating that schizophrenia patients and controls perform equally in recognizing whole-body expressions as well as human and animal auditory vocalizations. This does not rule out the possibility, however, that they have more difficulty with audiovisual stimuli. The auditory information may be harder to ignore for the patients either because the focused-attention task requires them to shut out one input system, or because ignoring it is harder in the case of human sounds.

GENERAL DISCUSSION

In the first experiment, we tested recognition of static emotional body language in a group of schizophrenia patients, nonschizophrenic psychotic patients, and controls. The results show a general emotion recognition impairment in the schizophrenia group. The impairment is also present in the nonschizophrenic psychosis group, but to a lesser extent. The present study shows that the emotion-recognition difficulties in schizophrenia, which have been previously documented with studies using facial expressions (for a review, see Mandal et al., 1998), extend to the recognition of body language. This deficit is consistent with findings from neuroimaging studies, showing that brain structures involved in perceiving emotional body language (for reviews, see de Gelder, 2006; de Gelder et al., 2010) show abnormalities in schizophrenia. Perception of bodily expressions activates not only brain areas associated with emotion perception, but also areas involved in action representation (de Gelder et al., 2004; Grèzes et al., 2007; Pichon, de Gelder, & Grèzes, 2008), and both of these structures show abnormalities in schizophrenia (e.g., Bertrand et al., 2008; Gur et al., 2002; Michalopoulou et al., 2008; Phillips et al., 1999; Pinkham, Hopfinger, Pelphrey, Piven, & Penn, 2008a). Impaired recognition of body expressions in schizophrenia and nonschizophrenic psychosis might have its roots in a dysfunction of the brain network involved in emotion perception, but possibly also in a deficit of the brain areas involved in action representation.

Recently, it has been suggested that the motor abilities of the observer are an important aspect of body language recognition. A link has been suggested between disorders with movement deficits and anomalous recognition of bodily expressions (de Gelder, 2006; de Gelder et al., 2008). It is possible that the motor problems associated with schizophrenia, such as catatonia, play an important role in recognizing emotional body language. Static images contain less information than dynamic stimuli and require the brain to generate and fill in the movement information.

The second experiment focused on multisensory integration of dynamic emotional body language, on the one hand, and both human and animal vocalizations on the other hand. The data show an increased integration of both modalities in the schizophrenia group, but only when the auditory information consists of human voices. These findings are compatible with an auditory-dominance hypothesis in schizophrenia, as a previous study with face–voice combinations showed a reduced influence of the facial expression on recognition of the vocal expression in schizophrenia patients (de Jong et al., 2009; Ross et al., 2007). A critical evaluation of this hypothesis in future studies would include task manipulation and instructions to categorize the emotion expressed by the auditory stimulus.

At the neuroanatomical level, binding of emotional information in the face and voice has been associated with activity in the amygdala (Dolan, Morris, & de Gelder, 2001; Ethofer et al., 2006). Interestingly, abnormal amygdala activity has been reported in schizophrenia in response to facial expressions (e.g., Gur et al., 2002; Michalopoulou et al., 2008; Phillips et al., 1999), and it is therefore likely that the anomalous multisensory integration also results from abnormal amygdalar activity.

The patients show a general impairment in recognizing static emotional body expressions, whereas there is no significant difference between both groups in recognizing dynamic, ecologically valid, whole-body expressions. A possible explanation for this finding concerns the cognitive task demands. Recognizing isolated static expressions requires a higher flexibility in order to compensate for the lack of information, as in direction of movement and speed of movement. Future research is needed to identify the specific processes that are impaired in schizophrenia patients when recognizing affective stimuli, possibly in relation to ecological validity.

REFERENCES

APA (2000). *Diagnostic and statistical manual of mental disorders* (DSM-IV-TR). Washington, DC: American Psychiatric Association.

Argyle, M. (1988). *Bodily communication*. London: Methuen.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality & Social Psychology*, *70*(3), 614–636.

Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 151–177). Oxford, UK: Oxford University Press.

Bertrand, M. C., Achim, A. M., Harvey, P. O., Sutton, H., Malla, A. K., & Lepage, M. (2008). Structural neural correlates of impairments in social cognition in first episode psychosis. *Social Neuroscience*, *3*(1), 79–88.

Borod, J. C., Martin, C. C., Alpert, M., Brozgold, A., & Welkowitz, J. (1993). Perception of facial emotion in schizophrenic and right brain-damaged patients. *Journal of Nervous and Mental Disease*, *181*(8), 494–502.

Borod, J. C., Pick, L. H., Hall, S., Sliwinski, M., Madigan, N., Obler, L. K., et al. (2000). Relationships among facial, prosodic, and lexical channels of emotional perceptual processing. *Cognition and Emotion*, *14*(2), 193–211.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, *9*, 353–356.

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, *7*(3), 242–249.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*(10), 460–467.

de Gelder, B., Bocker, K. B., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses. *Neuroscience Letters*, *260*(2), 133–136.

de Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjikhani, N. (2004). Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(47), 16701–16706.

de Gelder, B., & Van den Stock, J. (2011). Real faces, real emotions: Perceiving facial expressions in naturalistic contexts of voices, bodies and scenes. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 535–550). New York, NY: Oxford University Press.

de Gelder, B., Van den Stock, J., de Diego Balaguer, R., & Bachoud-Levi, A. C. (2008). Huntington's disease impairs recognition of angry and instrumental body language. *Neuropsychologia*, *46*(1), 369–373.

de Gelder, B., Van den Stock, J., Meeren, H. K., Sinke, C. B., Kret, M. E., & Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neuroscience and Biobehavioral Reviews*, *34*(4), 513–527.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14*(3), 289–311.

de Gelder, B., Vroomen, J., de Jong, S. J., Masthoff, E. D., Trompenaars, F. J., & Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenia. *Schizophrenics Research*, *72*(2–3), 195–203.

de Jong, J. J., Hodiamont, P. P., Van den Stock, J., & de Gelder, B. (2009). Audiovisual emotion recognition in schizophrenia: Reduced integration of facial and vocal affect. *Schizophrenia Research*, *107*(2–3), 286–293.

Dolan, R. J., Morris, J. S., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(17), 10006–10010.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., et al. (2006). Impact of voice on emotional judgment of faces: An event-related fMRI study. *Human Brain Mapping*, *27*(9), 707–714.

Feinberg, T. E., Rifkin, A., Schaffer, C., & Walker, E. (1986). Facial discrimination and emotional recognition in schizophrenia and affective disorders. *Archives of General Psychiatry*, *43*(3), 276–279.

Grèzes, J., Pichon, S., & de Gelder, B. (2007). Perceiving fear in dynamic body expressions. *NeuroImage*, *35*(2), 959–967.

Gur, R. E., McGrath, C., Chan, R. M., Schroeder, L., Turner, T., Turetsky, B. I., et al. (2002). An fMRI study of facial emotion processing in patients with schizophrenia. *American Journal of Psychiatry*, *159*(12), 1992–1999.

Hadjikhani, N., & de Gelder, B. (2003). Seeing fearful body expressions activates the fusiform cortex and amygdala. *Current Biology*, *13*(24), 2201–2205.

Hadjikhani, N., Joseph, R. M., Manoach, D. S., Naik, P., Snyder, J., Dominick, K., et al. (2009). Body expressions of emotion do not trigger fear contagion in autism spectrum disorder. *Social, Cognitive and Affective Neuroscience*, *4*(1), 70–78.

Heimberg, C., Gur, R. E., Erwin, R. J., Shtasel, D. L., & Gur, R. C. (1992). Facial emotion discrimination. III. Behavioral findings in schizophrenia. *Psychiatry Research*, *42*(3), 253–265.

Johnson-Frey, S. H., Maloof, F. R., Newman-Norlund, R., Farrer, C., Inati, S., & Grafton, S. T. (2003). Actions or hand–object interactions? Human inferior frontal cortex and action observation. *Neuron*, *39*(6), 1053–1058.

Joshua, N., & Rossell, S. (2009). Configural face processing in schizophrenia. *Schizophrenia Research*, *112*(1–3), 99–103.

Kee, K. S., Horan, W. P., Wynn, J. K., Mintz, J., & Green, M. F. (2006). An analysis of categorical perception of facial emotion in schizophrenia. *Schizophrenia Research*, *87*(1–3), 228–237.

Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., et al. (2003). Facial emotion recognition in schizophrenia: Intensity effects and error pattern. *American Journal of Psychiatry*, *160*(10), 1768–1774.

Mandal, M. K., Pandey, R., & Prasad, A. B. (1998). Facial expressions of emotions and schizophrenia: A review. *Schizophrenia Bulletin*, *24*(3), 399–412.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin and Review*, *3*, 215–221.

Michalopoulou, P. G., Surguladze, S., Morley, L. A., Giampietro, V. P., Murray, R. M., & Shergill, S. S. (2008). Facial fear processing and psychotic symptoms in schizophrenia: Functional magnetic resonance imaging study. *British Journal of Psychiatry*, *192*(3), 191–196.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*(8), 636–648.

Phillips, M. L., Williams, L., Senior, C., Bullmore, E. T., Brammer, M. J., Andrew, C., et al. (1999). A differential neural response to threatening and non-threatening negative facial expressions in paranoid and non-paranoid schizophrenics. *Psychiatry Research*, *92*(1), 11–31.

Pichon, S., de Gelder, B., & Grèzes, J. (2008). Emotional modulation of visual and motor areas by still and dynamic body expressions of anger. *Social Neuroscience*, *3*(3–4), 199–212.

Pinkham, A. E., Hopfinger, J. B., Pelphrey, K. A., Piven, J., & Penn, D. L. (2008a). Neural bases for impaired social cognition in schizophrenia and autism spectrum disorders. *Schizophrenia Research*, *99*(1–3), 164–175.

Pinkham, A. E., Hopfinger, J. B., Ruparel, K., & Penn, D. L. (2008b). An investigation of the relationship between activation of a social cognitive neural network and social functioning. *Schizophrenia Bulletin*, *34*(4), 688–697.

Pomarol-Clotet, E., Hynes, F., Ashwin, C., Bullmore, E. T., McKenna, P. J., & Laws, K. R. (2010). Facial emotion processing in schizophrenia: A non-specific neuropsychological deficit? *Psychological Medicine*, *40*(6), 911–919.

Pourtois, G., Debatisse, D., Despland, P. A., & de Gelder, B. (2002). Facial expressions modulate the time course of long latency auditory brain potentials. *Cognitive Brain Research*, *14*(1), 99–105.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Molholm, S., Javitt, D. C., & Foxe, J. J. (2007). Impaired multisensory processing in schizophrenia: Deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophrenia Research*, *97*(1–3), 173–183.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, *15*(2), 123–148.

Sprengelmeyer, R., Young, A. W., Schroeder, U., Grossenbacher, P. G., Federlein, J., Buttner, T., et al. (1999). Knowing no fear. *Proceedings of the Royal Society. Series B: Biological Sciences*, *266*(1437), 2451–2456.

Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: An ERP study on the body-inversion effect. *Neuroreport*, *15*(5), 777–780.

Tamietto, M., Geminiani, G., Genero, R., & de Gelder, B. (2007). Seeing fearful body language overcomes attentional deficits in patients with neglect. *Journal of Cognitive Neuroscience*, *19*(3), 445–454. doi: 10.1162/jocn.2007.19.3.445.

Van den Stock, J., Grèzes, J., & de Gelder, B. (2008a). Human and animal sounds influence recognition of body language. *Brain Research*, *1242*, 185–190.

Van den Stock, J., Peretz, I., Grèzes, J., & de Gelder, B. (2009). Instrumental music influences recognition of emotional body language. *Brain Topography*, *21*(3–4), 216–220.

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, *7*(3), 487–494.

Van den Stock, J., van de Riet, W. A., Righart, R., & de Gelder, B. (2008b). Neural correlates of perceiving emotional faces and bodies in developmental prosopagnosia: An event-related fMRI-study. *PLoS One*, *3*(9), e3195.

Vroomen, J., Driver, J., & de Gelder, B. (2001). Is crossmodal integration of emotional expressions independent of attentional resources? *Cognitive, Affective, & Behavioral Neuroscience*, *1*(4), 382–387.

Wolwer, W., Streit, M., Polzer, U., & Gaebel, W. (1996). Facial affect recognition in the course of schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience*, *246*(3), 165–170.