Multimodal Emotion Integration in Bipolar Disorder: An Investigation of Involuntary Cross-Modal Influences between Facial and Prosodic Channels

Tamsyn E. Van Rheenen,[1,2] AND Susan L. Rossell[1,2]

[1]Brain and Psychological Sciences Research Centre, Swinburne University
[2]Monash Alfred Psychiatry Research Centre, Central Clinical School, Monash University and the Alfred Hospital

Abstract

The ability to integrate information from different sensory channels is a vital process that serves to facilitate perceptual decoding in times of unimodal ambiguity. Despite its relevance to psychosocial functioning, multimodal integration of emotional information across facial and prosodic modes has not been addressed in bipolar disorder (BD). In light of this paucity of research we investigated multimodal processing in a BD cohort using a focused attention paradigm. Fifty BD patients and 52 healthy controls completed a task assessing the cross-modal influence of emotional prosody on facial emotion recognition across congruent and incongruent facial and prosodic conditions, where attention was directed to the facial channel. There were no differences in multi-modal integration between groups at the level of accuracy, but differences were evident at the level of response time; emotional prosody biased facial recognition latencies in the control group only, where a fourfold increase in response times was evident between congruent and incongruent conditions relative to patients. The results of this study indicate that the automatic process of integrating multimodal information from facial and prosodic sensory channels is delayed in BD. Given that interpersonal communication usually occurs in real time, these results have implications for social functioning in the disorder. (*JINS*, 2014, *20*, 525–533)

Keywords: Cognition, Affect, Mood disorders, Multisensory emotion processing, Unimodal processing, Audio-visual perception, Affective prosody, Facial emotion, Neuropsychology

INTRODUCTION

Bipolar disorder (BD) is a complex mood disorder associated with diminished quality of life and global functioning (Coryell et al., 1993; Cramer, Torgersen, & Kringlen, 2010; Freeman et al., 2009; Goodwin & Jamison, 1990; MacQueen, Young, & Joffe, 2001; Morriss et al., 2007; Saarni et al., 2010). A growing body of research suggests that social cognition is impaired in BD, with recent meta-analytic effect size estimates for overall facial emotion perception falling almost a quarter of a standard deviation below the healthy population mean (Rossell, Van Rheenen, Groot, Gogos, & Joshua, 2013; Samamé, Martino, & Strejilevich, 2012; Van Rheenen & Rossell, 2014). Although small, these impairments are nonetheless significant, and appear to contribute to

psychosocial dysfunction in the disorder (Hoertnagl et al., 2011; Martino, Strejilevich, Fassi, Marengo, & Igoa, 2011; Ryan et al., 2013). However, the predominant focus on the visual emotion processing domain (using facial expressions) in BD has hampered understandings of other social cognitive processes at play. Indeed, there is only preliminary research referencing prosodic processing or multimodal emotion integration in the disorder, despite the potential for deficits in these processes to detrimentally influence psychosocial outcome as well (Rossell et al., 2013; Van Rheenen & Rossell, 2013a, 2013b; Vederman et al., 2012).

In the natural environment, the stimulation of several sensory modalities occurs simultaneously, and the cognitive mechanisms underpinning the processing of information from these sources are thought to be strongly related (Borod et al., 2000; de Gelder & Jean, 2000). Although perception in the context of a single modality is sufficient in some contexts, the integration of equivocal (also referred to as redundant) information from these different sensory channels has been

Correspondence and reprint requests to: Tamsyn Van Rheenen, Monash Alfred Psychiatry research centre (MAPrc), Level 4, 607 St Kilda Road, Melbourne, VIC 3004, Australia. E-mail: tvanrheenen@swin.edu.au

found to augment meaningful and holistic perceptual decoding by improving both accuracy and speed of judgment (de Gelder et al., 2006; Paulmann & Pell, 2011; Pell, 2005). This multimodal integration reflects cross-modal influences between sensory channels that is thought to occur early in the time course of perception, where it serves to enrich perception, compensate for conflicts in cross-modal sensation and facilitate perceptual decoding in times of unimodal ambiguity (Alais & Burr, 2004; De Gelder & Bertelson, 2003; de Gelder et al., 2006; de Gelder, Pourtois, Vroomen, & Bachoud-Lévi, 2000; Vroomen, Driver, & Gelder, 2001). Indeed, auditory stimuli have been found to modulate visual perception and vice versa, with incompatibility between facial and speech information distorting perceptions (de Jong, Hodiamont, Van den Stock, & de Gelder, 2009; Kim, Seitz, & Shams, 2008; McGurk & MacDonald, 1976; Paulmann, Titone, & Pell, 2012; Shams, Kamitani, & Shimojo, 2004; Vroomen & De Gelder, 2000). As such, involuntary multimodal integration plays a substantial part in facilitating understandings of the world in a non-segmented, inclusive manner, and is thus, important for social and interpersonal functioning.

Despite BD being characteristically associated with poor psychosocial outcomes including a reduced capacity for meaningful, long-term interpersonal relationships (Australian Bureau of Statistics, 2007; Blairy et al., 2004; Tsai, Lee, & Chen, 1999), poor social skills (Goldstein, Miklowitz, & Mullen, 2006), and difficulties in social activities (Morriss et al., 2007), the possibility that these outcomes may be partially underpinned by abnormalities in multi-modal integration has not yet received adequate attention in the BD literature. However, an investigation of multimodal emotion processing in the disorder is justified in light of evidence suggesting that it is subserved by neural structures that are implicated in the pathophysiology of BD, including the temporal lobe, amygdala, anterior cingulate, and prefrontal cortex (de Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999; Dolan, Morris, & de Gelder, 2001; Laurienti et al., 2003; Phillips, Drevets, Rauch, & Lane, 2003; Phillips, Ladouceur, & Drevets, 2008; Pourtois, de Gelder, Bol, & Crommelinck, 2005). This, coupled with the large literature indicating that emotion perception from facial, and to a lesser extent prosody is impaired in the disorder, *and* evidence suggesting that multimodal integration itself is deficient in the genetically and phenotypically related disorder schizophrenia (Craddock, O'Donovan, & Owen, 2006, 2005; de Gelder et al., 2005; Kohler, Walker, Martin, Healey, & Moberg, 2010; Rossell et al., 2013; Van Rheenen & Rossell, 2013a, 2013b, 2014), certainly provides a foundation from which it is reasonable to assume that abnormalities in cross-modal influences between different sensory modalities may be present in BD.

One method of investigating multimodal integration is to use a focused attention paradigm comparing responses between conditions in which facial and prosodic modalities are fed congruent and incongruent emotional information. In this paradigm, when participants are explicitly instructed to make judgments based on the inputs of a particular modality,

better performance for congruent relative to incongruent stimuli (commonly referred to as priming) would indicate automatic cross-modal interference and thus, mandatory multimodal integration. On the other hand, matched performance between the conditions would indicate the absence of cross-modal priming.

Here, we present a study referencing this paradigm in a large cohort of BD patients compared to controls. That is, in the interests of ascertaining the nature of inter-sensory processes occurring during the time course of emotional perception in the disorder, we specifically sought to determine the extent to which the evaluation of complex emotional inputs from an attended visual channel (facial expressions) were biased by the concurrent presentation of stimuli in an unattended auditory channel (emotional prosody). By comparing facial emotion recognition on the parameters of both accuracy, and its more sensitive counterpart response latency, we aimed to examine the level at which potential group differences become apparent. We hypothesized that healthy individuals would demonstrate multimodal integration by a disproportionate response pattern to congruent relative to incongruent audio-visual emotional information. Given the lack of prior research on this topic in the BD literature, the extent to which emotional prosody would influence facial emotion recognition in individuals with BD remained an open question; a degree of impairment compared to controls was predicted however, due to prior evidence of unimodal deficits in BD cohorts.

MATERIALS AND METHODS

This study was approved by the Alfred Hospital and Swinburne University Human Ethics Review Boards and abided by the Declaration of Helsinki. Written informed consent was obtained from each participant before the study began.

Participants

The clinical sample comprised 50 patients (17 male, 33 female) diagnosed as having DSM-IV-TR BD (BD I $n = 38$; BD II $n = 12$) using the Mini International Neuropsychiatric Interview (MINI: Sheehan et al., 1998). Patients were recruited *via* community support groups and general advertisements and were all out-patients. Current symptomology was assessed using the Young Mania Rating Scale (YMRS: Young, Biggs, Ziegler, & Meyer, 1978) and the Montgomery Asberg Depression Rating Scale (MADRS: Montgomery & Asberg, 1979); there were 16 depressed (defined as those that met strict criteria for MADRS scores > 8), 4 (hypo) manic (defined as those that met strict criteria for YMRS scores > 8), 12 mixed (defined as those that met strict criteria for YMRS *and* MADRS scores > 8) and 18 euthymic patients (defined as those that met strict criteria for YMRS and MADRS scores ≤ 8). Those with current psychosis, co-morbid psychotic disorders, significant visual and auditory impairments, neurological disorder and/or a history of substance/alcohol

abuse or dependence during the past six months were excluded. Thirty two patients were taking antipsychotics, 15 were taking antidepressants, 16 were taking mood stabilizers, and 10 were taking benzodiazepines.

An age- and gender-matched control sample of 52 healthy participants (20 male, 32 female) were recruited for comparison purposes by general advertisement and contacts of the authors. Using the MINI screen, no control participant had a current diagnosis or previous history of psychiatric illness (Axis I). An immediate family history of mood and psychiatric disorder, in addition to a personal history of neurological disorder, current or previous alcohol/substance dependence or abuse, visual impairments and current psychiatric medication use was exclusion criteria for all controls.

All participants were fluent in English, were between the ages of 18 and 65 years and had an estimated pre-morbid IQ as scored by the Wechsler Test Of Adult Reading (WTAR) of > 75.

Materials

A task designed by the authors was administered to assess emotional multimodal integration across visual and auditory modes (described below). The Brief Assessment of Cognition in Schizophrenia - Symbol Coding subtest (taken from the MATRICS consensus cognitive battery and described in detail by Nuechterlein & Green, 2006) was used to co-vary out processing speed as a potential confound in the response time analysis.

Visual and Auditory Stimuli

The visual stimuli was taken from the widely used and well validated Ekman and Friesen series known as the Pictures of Facial Affect (POFA: Ekman & Friesen, 1976). The stimuli comprised black and white photographs of faces free of jewelry, spectacles, make up and facial hair (female and male) and expressing the emotions happy, sad, fear, and neutral. The faces were cropped to an oval shape spanning the top of the forehead to the bottom of the chin and excluding any hair and the ears on either side of the face. The auditory prosodic emotion stimuli comprised a series of sentences with neutral content ("the windows are made of glass"), matched for length and spoken by actors (male and female) who were directed to express each of four emotional prosodic tones (happy, sad, fear, and neutral). The task was presented binaurally *via* noise reduction headphones on a 14-inch Lenovo laptop computer and was run through Presentation (Neurobehavioral Systems Inc, 2012). Written instructions, an example and a set of practice trials were provided to participants before its commencement.

Design and Procedure

The task required participants to recognize emotional faces while being presented with a series of paired facial (visual) and prosodic (auditory) stimuli portraying either congruent or incongruent happy, sad, fearful or neutral emotional expressions[1]. Participants were instructed to keep their eyes on the screen at all times and to label a target emotion expressed by the facial stimuli whilst mentally blocking out the irrelevant prosodic stimuli. Responses were made *via* a labeled keyboard button press, where accuracy and response time data was recorded by the computer from 200 milliseconds (ms) onward. The task comprised 48 randomized trials, with 24 presentations of congruent stimuli (six pairs each for happy, sad, fear, or neutral expressions with some pairs being presented twice) and 24 presentations of incongruent stimuli (pairs representing different combinations of emotion with some pairs being presented twice or three times). Each trial lasted 3500 ms (including an inter-stimulus interval of 1000 ms); the prosodic stimuli presentation length was approximately 2500 ms and the facial stimuli presentation length was 2000 ms. The onset of the facial emotion stimuli were delayed by 500 ms to ensure that their departure from the computer screen coincided with the departure of the prosodic stimuli (which were longer in duration). This design was necessitated by the need to combine a short visual presentation with a longer auditory utterance. The delay in the onset of facial stimuli was suitable given that emotional information is not usually present in the initial fragment of a sentence, but is rather aggregated over its time course (de Gelder et al., 2005). The entire task took approximately 4 min to complete.

Statistical Analysis

Demographic and clinical group differences were assessed *via* χ^2 or independent samples *t* tests. We used a repeated measures (2[condition: congruent and incongruent]*2[group: controls and BD]) design to ascertain group differences between conditions for both accuracy and response time data. *Post hoc* paired sample *t* tests split by group were used to follow up significant results. To better understand the effects of diagnostic status and medication on task performance, all analyses were rerun in the patient group comparing those diagnosed with BD I (*n* = 38) *versus* BD II (*n* = 12), and those currently on or off different classes of medication. The effect of current mood status was also considered, however, given that the sample size of some of the mood phase subgroups were too small for meaningful analysis, we collapsed the mixed and manic groups into one (resulting *n* = 16) and compared this group to patients meeting criteria for euthymia (*n* = 18) or depression (*n* = 16). Bivariate correlations were also conducted to examine the relationship between mean response times to congruent and incongruent stimuli and symptom severity on the YMRS and MADRS.

---

[1] All pairs were made up of facial and prosodic emotional expressions matched for gender to eliminate confounds on this basis (i.e., male faces matched with male voices and female faces matched with female voices to create 10 male and 14 female pairs for the congruent stimuli and 14 male and 10 female pairs for the incongruent stimuli).

RESULTS

Descriptive Analyses

There was no significant difference in age, gender, or pre-morbid IQ between the two groups (see Table 1).

Multi-modal Integration

Figures 1 and 2 present mean accuracy and response time scores, respectively, for the recognition of facial emotions as a function of congruent or incongruent prosody in both patients and controls

Accuracy (% correct)

There was a main effect of condition ($F[1,100] = 23.43$; $p < .001$), with all participants performing more accurately in the congruent condition, relative to the incongruent condition (congruent: $M = 98.41$; $SD = 3.47$; incongruent: $M = 95.01$; $SD = 6.76$; $d = -.63$). However, there was no effect of group (Control $M = 97.15$; $SD = 3.50$; BD $M = 96.25$; $SD = 4.55$; $d = -.22$; $F[1,100] = 1.27$; $p = .26$) and no interaction effect ($F[1,100] = .15$; $p = .70$).

Response time (ms)

There was a main effect of condition ($F[1,100] = 12.58$; $p < .01$) and a main effect of group ($F[1,100] = 9.39$; $p < .01$), with response recognition taking longer for incongruent relative to congruent trials (congruent $M = 1584.42$; $SD = 429.52$, incongruent $M = 1672.94$; $SD = 401.21$; $d = .21$), and patients performing worse than controls overall (Control $M = 1515.54$; $SD = 375.79$; BD $M = 1746.34$; $SD = 384.71$; $d = -.31$). A significant interaction effect was also present ($F[1,100] = 5.62$; $p \leq .02$); such that the effect of condition on response time differed across the groups. Follow up analyses revealed a significant cross-modal effect of prosody on latency for recognizing facial emotions in controls only, such that response times were shorter when emotion congruent, relative to emotion incongruent prosody, accompanied facial stimuli ($t[51] = -4.84$; $p < .01$). This effect (reflected by the average ms difference between
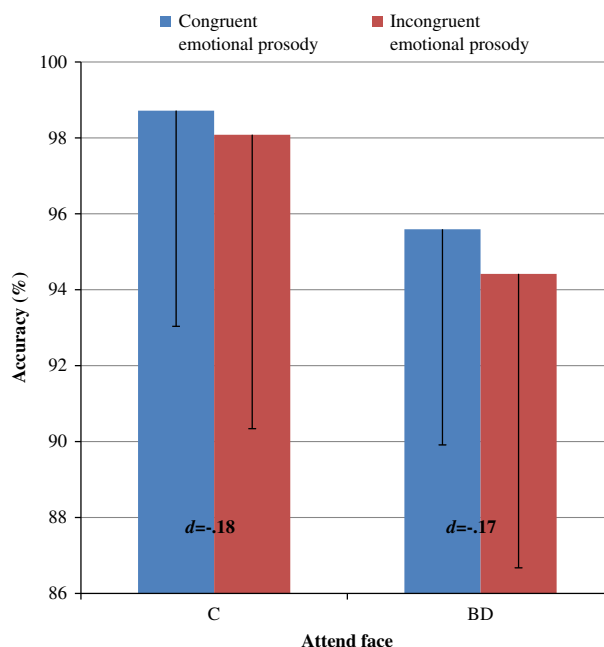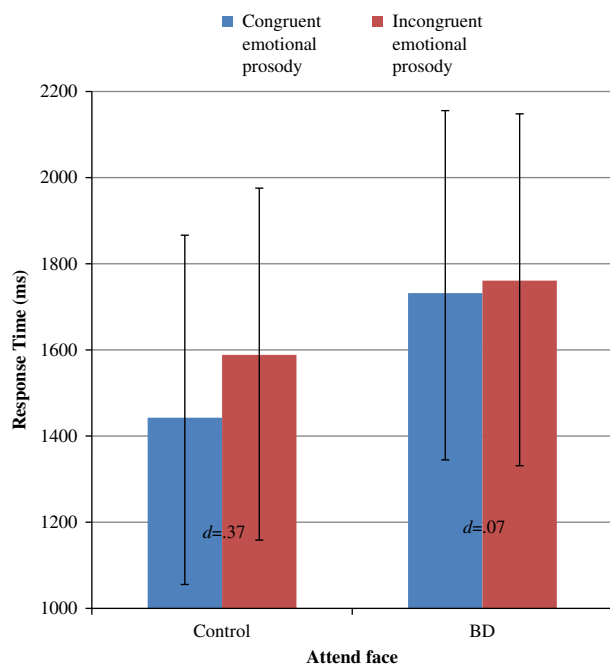

Fig. 1. Mean accuracy scores for congruent versus incongruent facial-prosodic emotional stimuli pairs in patients and controls: The pattern of performance for facial expression recognition according to congruent versus incongruent emotional prosody was the same across both groups. Cohen's d in bars represents the within group effect sizes for performance in congruent relative to incongruent conditions. Error bars represent standard deviations. Note that the effect size between groups for the congruent condition was $d = -.70$ and for the incongruent condition between groups $d = -.59$, with patients performing (non-significantly) worse than controls in both conditions.

responses in congruent and incongruent conditions) was four times greater for controls than it was for patients (Control $M = 145.75$; $SD = 217.10$; BD $M = 29.00$; $SD = 277.86$; *Cohen's d* $= -.47$). When the BACS-SC scores were entered into the repeated measures model to control for processing speed, the main effect of condition ($F[1,99] = 745.75$; $p = .88$) and group became non-significant ($F[1,99] = 2.43$; $p = .12$), but the significant interaction effect was preserved ($F[1,99] = 3.89$; $p \leq .05$).

Table 1. Demographic and clinical characteristics of the sample

Clinical variables		Controls		BD		Group comparisons		
		M	SD	M	SD	t/χ^2	df	p
Gender (M/F)	20/32			17/33		.22	1	.64
Age		33.98	14.28	38.46	13.02	-1.65	100	.10
WTAR		111.65	7.24	109.20	12.11	1.22	77.84	.23
MADRS		—	—	11.86	10.19	—	—	—
YMRS		—	—	6.31	5.50	—	—	—

Note: BD = bipolar disorder; M/F = Male / Female, WTAR = Wechsler Test of Adult Intelligence, YMRS = Young Mania Rating Scale, MADRS = Montgomery Asberg Depression Rating Scale.

Fig. 2. Mean response latencies for congruent versus incongruent facial-prosodic emotional stimuli pairs in patients and controls: Response times for facial expression identification in the BD group were less affected by incongruent emotional prosody than were response times in the control group. Cohen's d in bars represents the within group effect sizes for performance in congruent relative to incongruent conditions. Note that the effect size between groups for the congruent condition were $d = .71$ and for the incongruent condition $d = .44$, with patients taking longer than controls in both conditions. Error bars represent standard deviations.

Subgroup analyses

Table 2 presents descriptive statistics and correlations for the subgroup analyses. There were no between group main effects or interactions for patients diagnosed as having BD I versus BD II (all p's > .05), nor were there significant between group differences in the patient group for individuals on or off antipsychotics/anticonvulsants, antidepressants, lithium, and benzodiazepines (all p's > .05). There were also no between group main effects or interactions for patients classified as euthymic, depressed, or mixed/manic (all p's > .05). Further bivariate correlations supported this, indicating no associations between current depression (as measured by the MADRS) or mania (as measured by the YMRS) severity and response times to either congruent or incongruent stimuli.

DISCUSSION

This study is the first of its kind to investigate how the perceptual system of BD combines visual and auditory emotional information. By comparing the extent to which emotional prosody cross-modally influences facial expression recognition performance between BD and control cohorts, we were able to assess multimodal integration in the disorder. Our findings indicated that emotional prosody interfered with accuracy for recognizing facial expressions regardless of group status, with performance being better for the congruent relative to the incongruent condition. This occurred even though participants were instructed to attend only to information presented to the visual channel, which is consistent with previous literature suggesting that multimodal integration occurs pre-attentively (de Gelder & Jean, 2000; de Gelder, et al., 2005; Kim et al., 2008). Indeed, the less efficient processing of incongruent relative to congruent emotional information indicates that cues from different modalities were being integrated at a perceptual level.

In contrast, there were group differences in response times, with BD patients exhibiting difficulty in processing emotional facial expressions irrespective of the influence of congruent or incongruent prosodic information. This is largely supportive of prior work indicating that visual facial emotion impairments may represent a unique processing difficulty in BD (Vederman et al., 2012). An interaction effect was also evident, suggesting that it took longer for participants in the control group to recognize facial expressions when conflicted with incongruent emotional prosody. This effect was not apparent in the BD patients, with the absence of a priming/facilitation of emotion processing in this group relative to controls, signifying that sensory emotion integration was substantially impaired in this cohort. Indeed, the pattern of disproportionate latencies, in the absence of group differences in accuracy between conditions, suggests a subtle delay in automatic multimodal emotion integration in BD, such that typical redundancy based performance benefits are diminished. As this BD related impairment in implicitly extracting vocal emotional information was evident only on the sensitive response time parameter, a deficit in the normal automatic process of rapidly integrating information from different sensory sources appears to be quite subtle in nature.

This group difference is unlikely to be attributable to premorbid differences in processing speed, given that the group by condition interaction remained even when processing speed was statistically partialed out. As subgroup analyses did not reveal differences between patients diagnosed as having BD I versus BD II or meeting criteria for different symptomatic statuses, mood or diagnostic subtype are also unlikely to have had a significant impact. However, as these *post hoc* subgroup comparisons were underpowered, and as we were unable to directly compare these subgroups to controls due to the restricted sample size after stratification, the contribution of these factors cannot be completely ruled out.

There are other limitations to the study that should be considered when interpreting the results. First, the emotion integration task was newly developed in our lab and has not been validated in other clinical samples. Second, the presentation of the facial emotions over long intervals likely resulted in the ceiling level accuracy for the task.

Table 2. Descriptive statistics and correlations for subgroup comparisons in the BD group

| | | Accuracy (%) | | | | | Response time (ms) | | | | |
| | | BD I | | BD II | | | BD I | | BD II | | |
	Condition	M	SD	M	SD	d	M	SD	M	SD	d
Group differences	Congruent	98.14	4.18	97.92	3.77	.06	1738.60	394.58	1710.45	378.67	.07
between diagnostic	Incongruent	94.08	8.42	95.49	5.17	−.20	1736.21	453.98	1838.87	347.56	−.25
subtypes											
		Using		Not using			Using		Not using		
		M	SD	M	SD	d	M	SD	M	SD	d
Group differences	Antipsychotics-Congruent	98.44	2.95	97.22	6.03	.26	1801.59	389.01	1576.88	365.75	.60
between patients on/off	Antipsychotics-Incongruent	93.49	8.40	96.39	6.47	−.39	1786.79	474.07	1706.01	346.35	.19
different medications	Antidepressants-Congruent	98.33	2.64	97.92	4.74	.10	1880.93	426.69	1659.07	360.48	.56
	Antidepressants-Incongruent	91.11	11.45	96.00	5.02	−.55	1795.24	576.35	1744.96	360.86	.10
	Stabilisers-Congruent	98.18	3.39	98.00	4.55	0.04	1700.23	391.62	1745.17	398.25	−.04
	Stabilisers-Incongruent	94.01	11.07	94.62	5.82	−.07	1659.54	528.53	1813.38	377.41	−.33
	Benzodiazepines-Congruent	99.58	1.32	97.65	4.56	.57	1614.65	379.73	1761.01	394.88	−.38
	Benzodiazepines-Incongruent	95.42	7.20	94.14	8.13	.17	1682.07	353.64	1782.34	456.69	−.25
		MADRS	YMRS				MADRS	YMRS			
Pearson's correlations	Congruent	.14	.09				.04	−.06			
between conditions and	Incongruent	−.00	.11				.11	−.05			
mood severity											

Note: BD = bipolar disorder; MADRS = Montgomery Asberg Depression Rating Scale, YMRS = Young Mania Rating Scale

Although these intervals were necessary to accommodate the length of the auditory utterances, the timing of emotional facial stimuli in real-time interactions is much shorter. Thus, these results may not be ecologically valid. Third, there were too few stimuli in the current task design to reliably evaluate the effects of specific emotions. Subsequent research in the field would certainly do well to include more stimuli of each emotion type to investigate valence effects. Finally, the same face-emotion stimuli were repeated due to the limited number of stimuli in the POFA series. Thus, it is possible that our results are partly attributable to cross-contamination effects whereby responses on earlier trials affected responses on later trials using the same face and facial expression.

Despite these limitations, the current findings appear to indicate a level of hypo-integration in BD, at least in reference to the cross-modal influence of prosody on response latencies to facial expression recognition. Whether bidirectional cross-modal influences of facial expressions on emotional prosody is diminished in BD, however, remains to be seen. It is nonetheless likely that the typical magnitude of gains to be made on the basis of redundant multimodal information is diminished in these patients, at least at a subtle latency level. Further research directly comparing congruent multimodal emotion recognition to unimodal emotion recognition is needed to establish whether this is the case.

It also possible that the BD-related findings observed here are not specific to emotion, but rather relate to more general difficulties in receiving different information simultaneously. Indeed, it is widely recognized that patients with BD exhibit pervasive deficits in executive functioning that translate to difficulties in shifting between information sources and simultaneously thinking about multiple concepts (McKirdy et al., 2009; Melcher et al., 2013). Such neurocognitive deficits have been shown to underpin unimodal emotion processing in schizophrenia populations (Brekke, Kay, Lee, & Green, 2005), and it is certainly possible that our results reflect the downstream outcome of this cognitive inflexibility. Given that both executive and emotion perception impairments have recently been shown to predict occupational outcome in BD (Ryan et al., 2013), future work would certainly do well to establish a more coherent understanding of the interplay between cognitive and emotional processing abilities in this context.

As social communication is a complex process, our findings have substantial implications for the understanding of social functioning in BD. For example, given that multimodal integration is particularly necessary in times of unimodal ambiguity (Alais & Burr, 2004), *and* that facial emotion processing is impaired in BD, the fact that emotional prosody does not appear to implicitly guide the speed of perception of facial emotions in patients with the disorder may add to the significant psychosocial burden they carry. Indeed, this delay in generating a coherent emotional representation could certainly decrease cognitive efficiency in social situations where communication occurs in real time, in turn adversely influencing interpersonal behaviors. These findings certainly add to the existing research suggesting that emotion processing is important in facilitating healthy psychosocial outcomes (Hoertnagl et al., 2011; Martino et al., 2011; Ryan et al., 2013).

In conclusion, the results of this study indicate that the automatic process of rapidly integrating multimodal information from facial and prosodic sensory channels is impaired in BD. As we cannot make comment on whether this apparent multimodal integration impairment reflects an emotion specific phenomenon or more of a generalized audio-visual integration problem, prospective researchers would certainly be well placed to consider this when designing future multimodal investigations for the disorder.

ACKNOWLEDGMENTS

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262. doi:http://dx.doi.org/10.1016/j.cub.2004.01.029

Australian Bureau of Statistics (2007). *National survey of mental health and wellbeing: Summary of results*. Canberra: Australian Bureau of Statistics.

Blairy, S., Linotte, S., Souery, D., Papadimitriou, G.N., Dikeos, D., Lerer, B., … Mendlewicz, J. (2004). Social adjustment and self-esteem of bipolar patients: A multicentric study. *Journal of Affective Disorders*, *79*, 97–103.

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

Brekke, J., Kay, D.D., Lee, K.S., & Green, M.F. (2005). Biosocial pathways to functional outcome in schizophrenia. *Schizophrenia Research*, *80*, 213–225.

Coryell, W., Scheftner, W., Keller, M., Endicott, J., Maser, J., & Klerman, G.L. (1993). The enduring psychosocial consequences of mania and depression. *The American Journal of Psychiatry*, *150*, 720–727.

Craddock, N., O'Donovan, M.C., & Owen, M.J. (2006). Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. *Schizophrenia Bulletin*, *32*, 9–16. doi:10.1093/schbul/sbj033

Craddock, N., O'Donovan, M.C., & Owen, M.J. (2005). The genetics of schizophrenia and bipolar disorder: Dissecting psychosis. *Journal of Medical Genetics*, *42*, 193–204. doi:10.1136/jmg.2005.030718

Cramer, V., Torgersen, S., & Kringlen, E. (2010). Mood disorders and quality of life. A community study. *Nordic Journal of Psychiatry*, *64*, 58–62. doi:10.3109/08039480903287565

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*, 460–467. doi:http://dx.doi.org/10.1016/j.tics.2003.08.014

de Gelder, B., Böcker, K.B.E., 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*, 133–136. doi:http://dx.doi.org/10.1016/S0304-3940(98)00963-X

de Gelder, B., & Jean, V. (2000). The perception of emotions by ear and by eye. [Article]. *Cognition & Emotion*, *14*, 289–311. doi:10.1080/026999300378824

de Gelder, B., Meeren, H.K.M., Righart, R., Stock, J. v. d., van de Riet, W.A.C., & Tamietto, M. (2006). Beyond the face: Exploring rapid influences of context on face processing. In S. Martinez-Conde, S.L. Macknik, L.M. Martinez, J.M. Alonso & P.U. Tse (Eds.), *Progress in brain research* (Vol. 155, Part 2, pp. 37–48). Amsterdam: Elsevier.

de Gelder, B., Pourtois, G., Vroomen, J., & Bachoud-Lévi, A.-C. (2000). Covert Processing of faces in prosopagnosia is restricted to facial expressions: Evidence from cross-modal bias. *Brain and Cognition*, *44*, 425–444. doi:http://dx.doi.org/10.1006/brcg.1999.1203

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 schizophrenics. [3]. *Schizophrenia Research*, *72*, 195–203.

de Jong, J.J., Hodiamont, P.P.G., Van den Stock, J., & de Gelder, B. (2009). Audiovisual emotion recognition in schizophrenia: Reduced integration of facial and vocal affect. *Schizophrenia Research*, *107*, 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*, 10006–10010. doi:10.1073/pnas.171288598

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

Freeman, A.J., Youngstrom, E.A., Michalak, E., Siegel, R., Meyers, O.I., & Findling, R.L. (2009). Quality of life in pediatric bipolar disorder. *Pediatrics*, *123*, e446–e452. doi:10.1542/peds.2008-0841

Goldstein, T.R., Miklowitz, D.J., & Mullen, K.L. (2006). Social skills knowledge and performance among adolescents with bipolar disorder. *Bipolar Disorders*, *8*, 350–361. doi:10.1111/j.1399-5618.2006.00321.x

Goodwin, F.K., & Jamison, K.R. (1990). *Manic depressive illness*. New York: Oxford University Press.

Hoertnagl, C.M., Muehlbacher, M., Biedermann, F., Yalcin, N., Baumgartner, S., Schwitzer, G., … Hofer, A. (2011). Facial emotion recognition and its relationship to subjective and functional outcomes in remitted patients with bipolar I disorder. *Bipolar Disorders*, *13*, 537–544. doi:10.1111/j.1399-5618.2011.00947.x

Kim, R.S., Seitz, A.R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One*, *3*, e1532. doi:10.1371/journal.pone.0001532

Kohler, C.G., Walker, J.B., Martin, E.A., Healey, K.M., & Moberg, P.J. (2010). Facial emotion perception in schizophrenia: A meta-analytic review. *Schizophrenia Bulletin*, *36*, 1009–1019. doi:10.1093/schbul/sbn192

Laurienti, P.J., Wallace, M.T., Maldjian, J.A., Susi, C.M., Stein, B.E., & Burdette, J.H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Human Brain Mapping*, *19*, 213–223. doi:10.1002/hbm.10112

MacQueen, G.M., Young, L.T., & Joffe, R.T. (2001). A review of psychosocial outcome in patients with bipolar disorder. *Acta Psychiatrica Scandinavica*, *103*, 163–170.

Martino, D.J., Strejilevich, S.A., Fassi, G., Marengo, E., & Igoa, A. (2011). Theory of mind and facial emotion recognition in euthymic bipolar I and bipolar II disorders. *Psychiatry Research*, *189*, 379–384. doi:10.1016/j.psychres.2011.04.033

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

McKirdy, J., Sussmann, J.E.D., Hall, J., Lawrie, S.M., Johnstone, E.C., & McIntosh, A.M. (2009). Set shifting and reversal learning in patients with bipolar disorder or schizophrenia. *Psychological Medicine*, *39*, 1289–1293. doi:doi:10.1017/S0033291708004935

Melcher, T., Wolter, S., Falck, S., Wild, E., Wild, F., Gruber, E., … Gruber, O. (2013). Common and disease-specific dysfunctions of brain systems underlying attentional and executive control in schizophrenia and bipolar disorder. *European Archives of Psychiatry and Clinical Neuroscience*, [Epub ahead of print]. doi:10.1007/s00406-013-0445-9

Montgomery, S.A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, *134*, 382–389.

Morriss, R., Scott, J., Paykel, E., Bentall, R., Hayhurst, H., & Johnson, T. (2007). Social adjustment based on reported behaviour in bipolar affective disorder. *Bipolar Disorders*, *9*, 53–62. doi:10.1111/j.1399-5618.2007.00343.x

Neurobehavioral Systems Inc. (2012). Presentation,Version 14.8.

Nuechterlein, K., & Green, M.F. (2006). *MATRICS Consensus Cognitive Battery manual*. Los Angeles: MATRICS Assessment Inc.

Paulmann, S., & Pell, M.D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, *35*, 192–201.

Paulmann, S., Titone, D., & Pell, M.D. (2012). How emotional prosody guides your way: Evidence from eye movements. *Speech Communication*, *54*, 92–107. doi:10.1016/j.specom.2011.07.004

Pell, M. (2005). Nonverbal emotion priming: Evidence from the 'facial affect decision task'. *Journal of Nonverbal Behavior*, *29*, 45–73. doi:10.1007/s10919-004-0889-8

Phillips, M.L., Drevets, W.C., Rauch, S.L., & Lane, R. (2003). Neurobiology of emotion perception II: Implications for major psychiatric disorders. *Biological Psychiatry*, *54*, 515–528.

Phillips, M.L., Ladouceur, C.D., & Drevets, W.C. (2008). A neural model of voluntary and automatic emotion regulation: Implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Molecular Psychiatry*, *13*, 833–857.

Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, *41*, 49–59. doi:http://dx.doi.org/10.1016/S0010-9452(08)70177-1

Rossell, S.L., Van Rheenen, T.E., Groot, C., Gogos, A., & Joshua, N.R. (2013). Investigating affective prosody in psychosis: A study using the Comprehensive Affective Testing System. *Psychiatry Research*, *210*, 896–900. doi:10.1016/j.psychres.2013.07.037

Ryan, K.A., Vederman, A.C., Kamali, M., Marshall, D., Weldon, A.L., McInnis, M.G., & Langenecker, S.A. (2013). Emotion perception and executive functioning predict work status in euthymic bipolar disorder. *Psychiatry Research*, *210*, 472–478. doi:http://dx.doi.org/10.1016/j.psychres.2013.06.031

Saarni, S.I., Viertiö, S., Perälä, J., Koskinen, S., Lönnqvist, J., & Suvisaari, J. (2010). Quality of life of people with schizophrenia, bipolar disorder and other psychotic disorders. *The British Journal of Psychiatry*, *197*, 386–394. doi:10.1192/bjp.bp.109.076489

Samamé, C., Martino, D.J., & Strejilevich, S.A. (2012). Social cognition in euthymic bipolar disorder: Systematic review and meta-analytic approach. *Acta Psychiatrica Scandinavica*, *125*, 266–280. doi:10.1111/j.1600-0447.2011.01808.x

Shams, L., Kamitani, Y., & Shimojo, S. (2004). Modulations of visual perception by sound. In G.A. Calvert, C. Spence & B.E. Stein (Eds.), *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Sheehan, D.V., Lecrubier, Y., Harnett Sheehan, K., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G.C. (1998). The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59*, 22–33.

Tsai, S.-Y., Lee, J.-C., & Chen, C.-C. (1999). Characteristics and psychosocial problems of patients with bipolar disorder at high risk for suicide attempt. *Journal of Affective Disorders*, *52*, 145–152. doi:10.1016/s0165-0327(98)00066-4

Van Rheenen, T.E., & Rossell, S.L. (2013a). Auditory-prosodic processing in bipolar disorder; from sensory perception to emotion. *Journal of Affective Disorders*, *151*, 1102–1107.

Van Rheenen, T.E., & Rossell, S.L. (2013b). Is the non-verbal behavioural emotion-processing profile of bipolar disorder impaired? A critical review. *Acta Psychiatrica Scandinavica*, *128*, 163–178. doi:10.1111/acps.12125

Van Rheenen, T.E., & Rossell, S.L. (2014). Let's face it: Facial emotion processing is impaired in bipolar disorder. *Journal of the International Neuropsychological Society*, *20*, 200–208. doi:10.1017/S1355617713001367

Vederman, A.C., Weisenbach, S.L., Rapport, L.J., Leon, H.M., Haase, B.D., Franti, L.M., … McInnis, M.G. (2012). Modality-specific alterations in the perception of emotional stimuli in bipolar disorder compared to healthy controls and major depressive disorder. *Cortex*, *48*, 1027–1034.

Vroomen, J., & De Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1583–1590.

Vroomen, J., Driver, J., & Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective, & Behavioral Neuroscience*, *1*, 382–387. doi:10.3758/cabn.1.4.382

Young, R., Biggs, J., Ziegler, V., & Meyer, D. (1978). A rating scale for mania: Reliability, validity and sensitivity. *The British Journal of Psychiatry*, *133*, 429–435. doi: 10.1192/bjp.133.5.429