Cognition and Emotion

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

Involuntary processing of social dominance cues from bimodal face-voice displays

Virginie Peschard, Pierre Philippot & Eva Gilboa-Schechtman

COGNITION AND EMOTION, 2016
http://dx.doi.org/10.1080/02699931.2016.1266304

Routledge
Taylor & Francis Group

Involuntary processing of social dominance cues from bimodal face-voice displays

Virginie Peschard[a,b], Pierre Philippot[b] and Eva Gilboa-Schechtman[a]

[a]Psychology Department and the Gonda Brain Science Center, Bar Ilan University, Ramat Gan, Israel; [b]Laboratory for Experimental Psychopathology, Psychological Sciences Research Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

ABSTRACT
Social-rank cues communicate social status or social power within and between groups. Information about social-rank is fluently processed in both visual and auditory modalities. So far, the investigation on the processing of social-rank cues has been limited to studies in which information from a single modality was assessed or manipulated. Yet, in everyday communication, multiple information channels are used to express and understand social-rank. We sought to examine the (in)voluntary nature of processing of facial and vocal signals of social-rank using a cross-modal Stroop task. In two experiments, participants were presented with face-voice pairs that were either congruent or incongruent in social-rank (i.e. social dominance). Participants' task was to label face social dominance while ignoring the voice, or label voice social dominance while ignoring the face. In both experiments, we found that face-voice incongruent stimuli were processed more slowly and less accurately than were the congruent stimuli in the face-attend and the voice-attend tasks, exhibiting classical Stroop-like effects. These findings are consistent with the functioning of a social-rank bio-behavioural system which consistently and automatically monitors one's social standing in relation to others and uses that information to guide behaviour.

ARTICLE HISTORY
Received 10 November 2015
Revised 12 November 2016
Accepted 21 November 2016

KEYWORDS
Social rank; social dominance; Stroop; involuntary processing; facial and vocal processing

Social-rank cues communicate social status or social power within and between groups. Individuals high in social-rank tend to influence others and obtain prioritised access to valued resources, such as food and mates (e.g. Mazur, 2005). The ability to appropriately adjust behaviours based on the evaluation of conspecifics' social-rank is adaptive because it facilitates social coordination and minimises conflict, and as a result, maximises individual fitness (Cheng & Tracy, 2014). Since understanding social rank is so crucial to survival, several authors (e.g. Fournier, Moskowitz, & Zuroff, 2002; Johnson, Leedom, & Muhtadie, 2012; Zuroff, Fournier, Patall, & Leybman, 2010) suggest that humans and other species are equipped with a bio-behavioural system designed to rapidly detect social-rank related differences and to adjust interpersonal behaviour accordingly. This notion is supported by neuroimaging evidence implicating specific brain networks (e.g. limbic, prefrontal and intraparietal pathways) and of neurotransmitter systems (e.g. testosterone, dopamine) in the perception of social-rank cues and in the formation of social hierarchy (for a review see Watanabe & Yamamoto, 2015). Such social-rank bio-behavioural system appears to emerge early in development (e.g. Pun, Birch, & Baron, 2016; Thomsen, Frankenhuis, Ingold-Smith, & Carey, 2011) and to operate fluently in adults (Zitek & Tiedens, 2012).

A growing body of behavioural studies has documented that social-rank information is readily processed from visual cues. For instance, using a variant of the affective Simon paradigm, Moors and De Houwer (2005) showed that participants automatically appraised the dominant (high social-rank) or submissive (low social-rank) status based on pictures depicting scenes of two interacting partners while making judgments on the basis of their spatial positions. Relatedly, eye-tracking studies (e.g. DeWall & Maner, 2008;

CONTACT Virginie Peschard peschardv@gmail.com

Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010; Maner, DeWall, & Gailliot, 2008) demonstrated that individuals displaying high dominance characteristics attract more attention than did individuals displaying low dominance characteristics. Furthermore, several gaze-cueing investigations found that participants were more likely to follow the gaze of high rather than low status faces (Dalmaso, Galfano, Coricelli, & Castelli, 2014; Dalmaso, Pavan, Castelli, & Galfano, 2012). Finally, Ratcliff, Hugenberg, Shriver, and Bernstein (2011) extended these findings by showing that high status targets are not only preferentially attended to, but also better remembered than low status targets. Taken together, these results suggest that visual social-rank cues attract attention, are effortlessly processed and preferentially remembered. In the aforementioned studies, social-rank of the targets was experimentally manipulated by varying their dress (DeWall & Maner, 2008; Maner et al., 2008; Ratcliff et al., 2011) and the perceived social dynamics between targets (Moors & De Houwer, 2005), as well as by using textually presented information about professional background (Dalmaso et al., 2012; Dalmaso et al., 2014; Ratcliff et al., 2011).

Social-rank can also be readily inferred from physical characteristics of individuals, such as their facial features (e.g. Oosterhof & Todorov, 2008; Sutherland et al., 2013). For example, facial masculinity has been found to positively correlate with impressions of social dominance (Jones et al., 2010; Main, Jones, DeBruine, & Little, 2009). Similarly, facial maturity was linked to higher perceived social dominance (Todorov, Said, Engell, & Oosterhof, 2008). Moreover, individuals whose neutral facial expression resembles certain emotional expressions (e.g. facial features associated with anger) are more likely to be perceived as possessing traits associated with those emotions (e.g. dominance for anger) (Zebrowitz, 1996, 1997; for a review of overgeneralisation effects see Zebrowitz & Montepare, 2015). Dominance and competence judgements from faces are made as early as 200 ms after stimulus presentation (Chiao et al., 2008; Santamaría-García, Burgaleta, & Sebastián-Gallés, 2015) and can affect a variety of social outcomes (Mueller & Mazur, 1996; Sussman, Petkova, & Todorov, 2013).

Social-rank is conveyed by vocal features in addition to visual ones. Specifically, voice pitch (as determined by fundamental frequency) has been identified as an efficient and rapid transmitter of social dominance (McAleer, Todorov, & Belin, 2014). Several empirical studies have shown that vocal masculinity (i.e. low-pitched voice) is associated with impressions of greater social dominance (e.g. Puts, Gaulin, & Verdolini, 2006; Wolff & Puts, 2010). A recent investigation (Cheng, Tracy, Ho, & Henrich, 2016) has indicated that individuals can strategically modulate their vocal pitch in face-to-face rank contests and that these modulations predict listeners' perceptions of social dominance and rank outcomes. Voice pitch has been shown to quickly influence social dominance judgments (McAleer et al., 2014) and to influence election of leaders (e.g. Klofstad, Anderson, & Nowicki, 2015; Klofstad, Anderson, & Peters, 2012) and mate preferences (e.g. Vukovic et al., 2011).

Taken together, this evidence suggests that information about social-rank is fluently processed in both visual and auditory modalities. However, so far, the investigation on the processing of social-rank has been limited to studies in which information from a single modality was assessed or manipulated (for an exception see Rezlescu et al., 2015). Yet, in everyday communication, multiple information channels are used to express and understand social-rank. Typically, high social-rank faces are simultaneously presented with high social-rank voices. Several lines of research provide support that different information modalities are combined into a unified percept, and this integration facilitates social comprehension. For example, in the emotion perception domain, empirical studies have shown that emotion categorisation is improved in response to emotionally congruent face-voice pairs as compared to unimodal stimulations (e.g. Collignon et al., 2008; Paulmann & Pell, 2011). By contrast, if faces and voices convey different emotional expressions (i.e. incongruent), they are categorised less accurately and more slowly than congruent pairs. Because such cross-modal effects arise despite the explicit instruction to focus on a single channel, interactions between facial and vocal emotional information have been suggested to occur independently of attentional allocation (Collignon et al., 2008; de Gelder & Vroomen, 2000).

We believe that social-rank research may gain from following the lead of emotion perception research. Specifically, a direct comparison of responses to congruent and incongruent bimodal stimuli while ignoring one of the two modalities can provide a useful method to assess the (in)voluntary nature of processing of facial and vocal signals of social-rank.

The present study

The aim of this study was to examine the (in)voluntary processing of social-rank facial and vocal cues. To address this question, we conducted two experiments using a cross-modal Stroop task. In the classical Stroop task (Stroop, 1935), participants are instructed to focus on the task-relevant dimension of a multidimensional stimulus while ignoring the irrelevant dimension. The stimulus is a colour word (e.g. "red" written in blue), the task-relevant dimension is the colour and the task-irrelevant dimension is the word's meaning. The Stroop effect refers to the phenomenon wherein performance of the target task is hampered (i.e. slower reaction times (RTs) and/or lower accuracy) when the word's name is incongruent relative to when it is congruent with the colour name (e.g. "red" written in red), despite the instruction not to read the words. The Stroop effect is typically interpreted as a hallmark failure of selective attention, indicating involuntarily attention to the extraneous dimension (for reviews, see MacLeod, 1991; Melara & Algom, 2003).

Cross-modal Stroop differs from the traditional Stroop task in that the task-relevant and task-irrelevant dimensions are presented in different modalities. Specifically, in the current study, participants were presented with face-voice pairs that were either congruent (e.g. dominant face and voice; submissive face and voice) or incongruent (e.g. submissive face and dominant voice; submissive voice and dominant face). They were asked to identify the social dominance of the face ignoring the voice (Face-attend), or the social dominance of the voice ignoring the face (Voice-attend). We predicted to observe Stroop-like effects in the processing of these cross-modal cues. Such effects were expected to be evidenced in slower RTs and lower accuracy in response to incongruent relative to congruent face-voice pairs.

Pilot experiment

Method

Participants

Thirty-four participants (22 women and 12 men; mean age = 24.76, SD = 4.54) took part in the study in exchange for payment (an equivalent of $5). Participants were recruited through advertisements in Bar-Ilan University. All were native Hebrew speakers.

Materials

The stimulus set was composed of neutral facial expressions combined with vocal utterances. The faces were selected from a standardised database (Oosterhof & Todorov, 2008), a collection of 25 Caucasian male faces generated using the Facegen Modeler programme (http://facegen.com) Version 3.1. Each individual face was adapted into seven exemplars varying along the dimension of dominance (from 0% dominant to 100% dominant) using the dominance computer model of Oosterhof and Todorov (2008). We randomly selected 10 different facial identities out of the 25 facial identities: 5 faces rated as 0% dominant and 5 faces rated as 100% dominant. All the pictures were resized to 180 × 220 pixels bitmap. One high-dominant face and one low-dominant face were used for practice trials.

For the vocal stimuli, 10 individuals were individually recorded (using Adobe Audacity 2.1) uttering the English word "Hi" in a high-dominant or low-dominant manner, resulting in a total of 10 recordings. The mean duration of the recording was 252 ms ± 83 ms. One high-dominant vocal utterance and one low-dominant vocal utterance were included in the practice trials. The voices were normalised to the same mean signal amplitude. Auditory stimuli were delivered through headphones and the sound volume was set at about 60 dB for all participants.

Two different types of stimulus pairs were created: Congruent (high-dominant face/high-dominant voice, low-dominant face/low-dominant voice), and incongruent (high-dominant face/ low-dominant voice, low-dominant face/high dominance voice). There was an equal number of congruent and incongruent stimuli. The faces and the voices started at the same time. The study contained 16 stimuli (8 models × 2 pairing types).

Procedure

Participants were greeted, and seated individually in a quiet, soundproof room. After signing an informed consent form, participants performed the Stroop task. The viewing distance was 50 cm. Two versions of the Stroop task were included: a face-attend task and a voice-attend task. Both tasks contained 12 practice trials followed by 64 (16 stimuli × 4 repetitions) experimental trials. The presentation of the stimuli within each task was randomised, and the order of the tasks was counterbalanced across participants. Instructions were provided on the computer screen.

Each trial began with the presentation of a white fixation cross displayed in the middle of a black coloured background for 500 ms. The fixation cross was followed by a face-voice pair for 800 ms. Immediately after the stimulus offset, a black screen was presented and remained on the screen until a response was recorded. *Participants were asked to* categorise the presented cues as "Dominant" or "Not dominant". In the face-attend task, participants had to make their decision on the basis of the facial information only, ignoring the vocal information, whereas in the voice-attend task, they had to base their judgment on the voice only, ignoring the facial information. In the latter task, the instructions emphasised that the participants had to look at the faces (e.g. "Don't close your eyes or look away from the face"). Responses were made by pressing one of two keys ("v" or "b") on a QWERTY keyboard with the participant's preferred hand. The inter-trial interval was 1000 ms.

Following the completion of the Stroop tasks, participants rated the dominance of face-voice pairs, as well as each face and voice separately using Qualtrics software (Qualtrics, Provo, UT). The unimodal blocks (faces only, voices only) were followed by the presentation of the bimodal block (combined faces and voices). The order of the unimodal blocks was counterbalanced across participants. In each block, a cue was presented at the centre of the screen with a response scale below. Participants were instructed to rate how dominant the person was on the basis of the presented cue. The response scale ranged from 0% (Not at all dominant) to 100% (Very dominant). At the end of the experiment, participants were debriefed, paid and thanked for their participation.

Results and discussion

Subjective face and voice ratings

To analyse the subjective ratings of the unimodal stimuli, we conducted a 2 (Modality: facial, vocal) × 2 (Dominance: high, low) repeated measures analysis of variance (ANOVA), with repeated measurement on the two factors. Table 1 presents the means and standard deviations (SDs) of dominance ratings. A significant main effect of Dominance was found $F(1,33) = 150.88$, $p < .001$, $\eta_p^2 = .82$, confirming that high-dominant stimuli were rated as more dominant (M = 71.30, SD = 13.62) than were low-dominant stimuli (M = 33.43, SD = 12.02). No main effect of Modality or Modality × Dominance interaction were identified (all Fs < 2.08).

We also analysed the subjective ratings of the bimodal stimuli using a 2 (Facial dominance: high, low) × 2 (Vocal dominance: high, low) repeated measures ANOVA. Facial dominance and Vocal dominance were defined as within-subject factors. Main effects of Facial dominance $F(1,33) = 40.90$, $p < .001$, $\eta_p^2 = .55$ and Vocal dominance $F(1,33) = 122.18$, $p < .001$, $\eta_p^2 = .79$ were identified, as well as a significant Facial dominance × Vocal dominance interaction $F(1,33) = 6.80$, $p < .05$, $\eta_p^2 = .17$. This two-way interaction revealed that the ratings were higher for congruent high-dominant pairs, followed by low-dominant face/high-dominant voice pairs, high-dominant face/low-dominant voice pairs and congruent low-dominant pairs ($p < .001$ for all paired t-tests).

Stroop task

Data reduction were performed in four successive steps. First, incorrect responses were discarded (13%). Second, RTs less than 150 ms and RTs more than 3 SDs from the individual mean RTs (<0.02%) were eliminated. Finally, one participant was discarded from the analyses due to low level of accuracy (<50%).

Afterwards, response accuracy and RTs were subjected to a 2 (Task: face-attend, voice-attend) × 2 (Congruence: congruent, incongruent) × 2 (Order: face-attend-first, voice-attend-first) repeated measures ANOVA with repeated measurement on the first two factors. Figure 1 presents the means and standard errors of the accuracy and RTs measures for the congruent and incongruent conditions.

Response accuracy. Consistent with our hypothesis, a significant main effect of Congruence was observed, $F(1,31) = 25.79$, $p = .001$, $\eta_p^2 = 45$, indicating greater accuracy in congruent as compared to incongruent pairs. The ANOVA also yielded a significant main effect of Task, $F(1,31) = 14.15$, $p < .001$, $\eta_p^2 = .31$, revealing less errors in the face-attend task (M = 90.72%, SD = 8.34) than in the voice-attend task (M = 82.20%, SD = 9.79). No other main effects or interactions were significant (all Fs < .89).

Reaction times. Consistent with our predictions, the ANOVA revealed a main effect of Congruence, $F(1,31) = 17.74$, $p < .001$, $\eta_p^2 = .36$, demonstrating longer response times for incongruent pairs than for congruent ones. In addition, a significant main effect of Task, $F(1,31) = 21.05$, $p < .001$, $\eta_p^2 = .40$, indicated that the face-attend task was performed

Table 1. Mean of dominance ratings with standard deviation in parentheses.

		Pilot experiment		Experiment	
		High dominant	Low dominant	High dominant	Low dominant
Unimodal	Face	73.39 (14.03)	33.61 (14.18)	72.76 (15.48)	29.05 (14.54)
	Voice	69.21 (15.17)	33.24 (13.96)	67.34 (15.58)	26.89 (13.60)
		Vocal information			
Bimodal		High dominant	Low dominant	High dominant	Low dominant
	Facial information				
	High dominant	72.60 (16.87)	45.86 (14.47)	71.54 (15.46)	44.32 (17.04)
	Low dominant	62.85 (15.19)	30.99 (14.26)	46.55 (15.41)	23.91 (13.66)

faster (*M* = 781 ms, SD = 195) than the voice-attend task (*M* = 943 ms, SD = 301). Furthermore, a significant Task × Order interaction $F(1,31) = 16.54$, $p < .001$, $\eta_p^2 = .35$ was found. To follow-up this interaction, analyses were conducted separately for each Order. When the voice-attend-was performed first, the response times in the first (voice-attend) task were slower than in the second (face-attend) task (*M* = 1034 ms, SD = 386 vs. *M* = 737 ms, SD = 214), $F(1,16) = 26.18$, $p < .001$, $\eta_p^2 = .62$. In contrast, when face-attend task was presented first, there were no significant differences in the RTs for the two tasks (*M* = 846 ms, SD = 122 for the voice-attend task vs. *M* = 827 ms, SD = 166 for the face-attend task), $F(1,15) = .27$, NS. This result may reflect the difference in task difficulty, such that decreasing task difficulty (from voice-attend to face-attend tasks) may improve RTs. No other significant main effects or interactions were found (all *F*s < 3.41).

Consistent with our hypotheses, social dominance information conveyed by the to-be-ignored channel interfered with the processing of social dominance information from the attended channel. Specifically, we found that face-voice incongruent stimuli were processed more slowly and less accurately than congruent stimuli in both the face-attended and the voice-attended tasks. These findings are in line with existing literature suggesting that signals of social-rank can be processed in a relatively automatic way (DeWall & Maner, 2008; Maner et al., 2008; Moors & De Houwer, 2005).

Several limitations need to be acknowledged. First, an important limitation of this Pilot Experiment is that facial and vocal stimuli differed in their presentation duration (i.e. 800 ms for faces, 250 ms for voices). The increased presentation time of faces might therefore explain the higher level of task difficulty in the voice-attend task in comparison with the face-attend task. Second, facial stimuli consisted of computer-generated faces, resulting in a relatively low ecologically relevant material. We sought to address these limitations in our next Experiment.

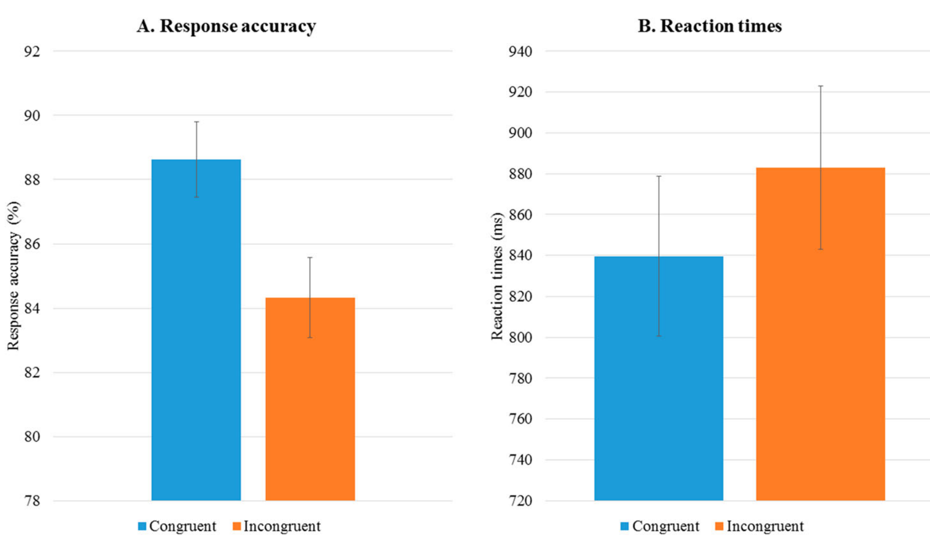

Figure 1. Means and standard errors of accuracy and RTs for the congruent and incongruent conditions.

Experiment

Method

Participants

Fifty-one participants (25 women and 26 men; mean age: 25.14, SD: 8.24) participated to the study in exchange for payment or academic credit. They were recruited through advertisements in Bar-Ilan University. All were native Hebrew speakers.

Materials

The facial stimuli were selected from the Chicago face database which consists of 158 faces of males and female faces of different ethnicity and ages (Ma, Correll, & Wittenbrink, 2015). Four different Caucasian identities were preselected based on their facial features of high and low dominance. The vocal stimuli were 4 native Hebrew male voices uttering the pseudo-word "Materet" in a high-dominant or low-dominant manner. A pretest conducted on 41 participants confirmed that high-dominant stimuli were rated as more dominant (Faces: $M = 60.93$, SD = 15.68; Voices: $M = 49.90$, SD = 18.91) than were the low-dominant stimuli (Faces: $M = 32.77$; SD = 18.38; Voices: $M = 37.65$; SD = 20.77), $F(1,40) = 80.50$, $p < .001$, $\eta_p^2 = .67$. The pictures were resized to 175 × 250 pixels bitmap. All vocal utterances were 690 ms ± 28 ms in duration and normalised to the same mean signal amplitude. In addition to the experimental stimulus set, one high-dominant face and one low-dominant face were chosen for practice trials.

As in the Pilot Study, two different types of stimulus pairs were created: Congruent (high-dominant face/ high-dominant voice, low-dominant face/low-dominant voice), and incongruent (high-dominant face/ low-dominant voice, low-dominant face/high dominance voice) with an equal number of congruent and incongruent cues. The face started at the same time as the voice. The study contained 16 stimuli: 4 (Individual faces) × 4 (Individual voices).

Procedure

The procedures of the experiments were similar in that the participants were presented with combined face-voice pairs, and their task was to ignore one modality (e.g. vocal) and base their judgment of a specific stimulus on the other modality only. Importantly, in this Experiment, the duration of the presentation of both modalities was 700 ms. This is in contrast to the Pilot Experiment in which the duration of the vocal utterances was about 250 ms while the visual stimuli were presented for 800 ms. Moreover, each trial began with the presentation of the response labels ("dominant" and "non-dominant") which remained on the screen for the duration of the trial (whereas in the Pilot Experiment, response labels were displayed on the keyboard). Third, the response keys ("e" or "i") were counterbalanced across participants. Fourth, the dominance rating task was performed a week after the completion of the Stroop task. Finally, these data were collected as part of a larger research.

Result and discussion

Subjective face and voice ratings

One participant was discarded from the ratings analyses (also from the Stroop task analyses) because of abnormally low variance in the ratings (i.e. rated most stimuli, 1) and one participant did not complete the dominance rating task a week after the completion of the Stroop task.

The subjective ratings of the unimodal stimuli were analysed by performing a 2 (Modality: facial, vocal) × 2 (Dominance: high, low) repeated measures ANOVA, with repeated measurement on the two factors (see Table 1). A significant main effect of Dominance was observed, $F(1,48) = 430.74$, $p < .001$, $\eta_p^2 = .90$, with high-dominant stimuli being rated as more dominant ($M = 70.05$, SD = 12.24) than low-dominant stimuli ($M = 27.97$, SD = 10.49). There was also a significant main effect of Modality, $F(1,48) = 4.11$, $p < .05$, $\eta_p^2 = .08$, showing that facial stimuli received higher ratings ($M = 50.90$, SD = 11.20) than vocal stimuli ($M = 47.11$, SD = 10.92). However, no Modality × Dominance interaction was identified, $F(1,48) = .69$, p = NS.

To analyse the subjective ratings of the bimodal stimuli, we conducted a 2 (Facial dominance: high, low) × 2 (Vocal dominance: high, low) repeated measures ANOVA with Facial dominance and Vocal dominance defined as within-subject factors. The analyses show significant main effects of Facial dominance $F(1,48) = 128.11$, $p < .001$, $\eta_p^2 = .73$ and Vocal dominance, $F(1,48) = 127.74$, $p < .001$, $\eta_p^2 = .73$, as well as a significant Facial dominance × Vocal dominance interaction, $F(1,48) = 8.36$, $p < .01$, $\eta_p^2 = .15$. The ratings were higher for congruent high-dominant pairs, followed by incongruent pairs, and congruent low-dominant pairs ($p < .001$, for all paired t-tests, except between incongruent pairs).

Stroop task

Data reduction followed the same algorithm as in the Pilot Experiment. First, incorrect responses were discarded (19.51%). Second, RTs less than 150 ms (<0.02%) or RTs more than 3 SDs from the individual mean RTs (<0.02%) were eliminated. Finally, four participants were discarded from the analyses due to low level of accuracy (≤50%).

Subsequently, a 2 (Task: face-attend, voice-attend) × 2 (Congruence: congruent, incongruent) × 2 (Order: face-attend-first, voice-attend-first) repeated measures ANOVA with repeated measurement on the first two factors was computed for both response accuracy and RT. Figure 2 presents the means and standard errors of the accuracy and RTs measures for the congruent and incongruent conditions.

Response accuracy. Consistent with our hypothesis and with the results of the Pilot Experiment, a significant main effect of Congruence was observed $F(1,45) = 4.30$, $p < .05$, $\eta_p^2 = .09$, demonstrating greater accuracy in congruent, as compared to incongruent pairs. No other main effects or interactions were found significant (all Fs < 2.33).

Reaction times. Consistent with our hypotheses and with the results of the Pilot Experiment, the ANOVA showed a significant main effect of Congruence $F(1,45) = 8.10$, $p < .01$, $\eta_p^2 = .15$, with slower responses for incongruent pairs than for congruent ones. In addition a significant main effect of Task $F(1,45) = 87.67$, $p < .001$, $\eta_p^2 = .66$ was found, revealing longer response times in the voice-attend task ($M = 812$ ms, SD = 175) than in the face-attend task ($M = 652$ ms, SD = 153). This main effect was further modulated by Order $F(1,45) = 9.38$, $p < .005$, $\eta_p^2 = .17$. To decompose the Task × Order interaction, separated ANOVAs for each Order were conducted. Results indicated a significant main effect of Task for both Orders (face-attend-first: $F(1,25) = 33.83$, $p < .001$, $\eta_p^2 = .57$; voice-attend-first order $F(1,20) = 48.82$, $p < .001$, $\eta_p^2 = .71$), such that participants were slower in the voice-attend task than in the face-attend task, regardless of task Order. However, in line with the Pilot Experiment, the difference in RTs between the voice-attend and face-attend tasks was larger when the voice-attend task was performed first ($M_{diff} = 221$ ms, $SD_{diff} = 144.41$) than when the face-attend task was performed first ($M_{diff} = 111$ ms, $SD_{diff} = 96.70$), $F(1,45) = 9.60$, $p < .005$. No other significant main effects or interactions were found (all Fs < 3.01).

Consistent with our hypotheses and the results of the Pilot Experiment, we found that social dominance information from the to-be-ignored channel compromised the processing of social dominance information from the attended channel. Additionally, like in the Pilot Experiment, our data indicated that participants were faster in the face-attend than in the voice-attend task, suggesting that it was easier to ignore the voices than the faces. In contrast to the Pilot Experiment, there was no significant effect of Task at the level of response accuracy.

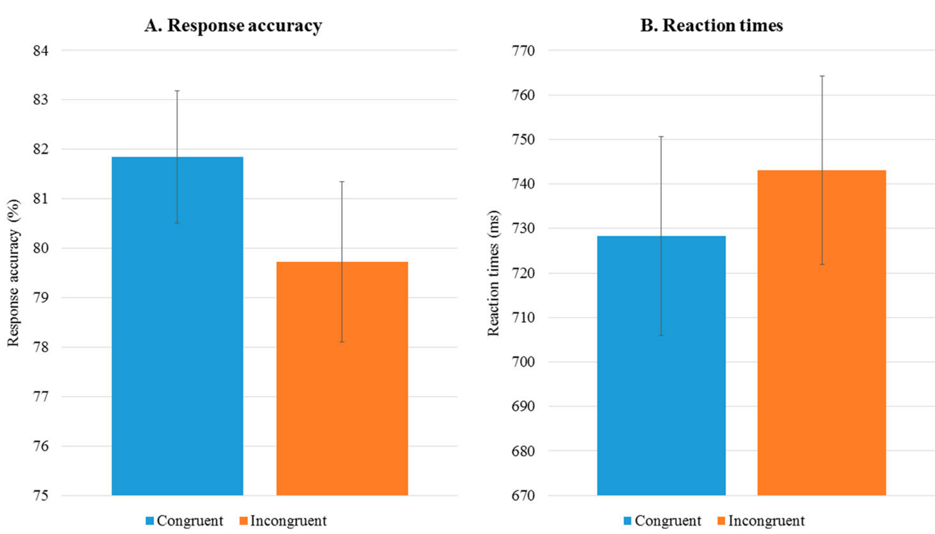

Figure 2. Means and standard errors of accuracy and RTs for the congruent and incongruent conditions.

Interestingly, the interference effects from the task-irrelevant channel were smaller in the Experiment than in the Pilot Experiment. To understand these effects, we compared the subjective ratings of the unimodal stimuli in the two experiments. These analyses did not reveal any significant effects or interactions involving voice-type (word in the Pilot Experiment, pseudo-word in the Experiment) and face-type (artificial in the Pilot Experiment, naturalistic in the Experiment). We also contrasted the subjective ratings of the different types of face-voice pairs in the two experiments. This analysis revealed lower ratings in the Experiment than in the Pilot Experiment for low-dominance face/high-dominance voice pairs ($F(1,81) = 25.90$, $p < .001$) and low-dominant face/low-dominant voice pairs $F(1,81) = 5.19$, $p < .05$), thus possibly decreasing the subjective distinction between congruent and incongruent stimuli.

General discussion

In two experiments, participants were presented with face-voice pairs that were either congruent or incongruent in social-rank. Specifically, they were requested to focus their attention to only one sensory modality at a time and *to* categorise the social dominance of the attended cue while disregarding the irrelevant modality. Both studies showed that social dominance information is difficult to ignore: slower and less accurate responses were observed when the dominance information of the to-be-ignored modality was incongruent with the dominance information of the attended modality. The fact that this effect was observed despite the clear instructions to attend to only one channel may suggest that the processing of social dominance is automatic to some extent (i.e. involuntary). These findings dovetail with the neuroimaging and electrophysiological findings (Chiao et al., 2008; Feng, Tian, Feng, & Luo, 2015; Santamaría-García et al., 2015), suggesting early cortical processing of social hierarchy information.

These results add to the existing literature on social-rank perception in several ways. First, using the gold standard method of assessing automatic attention – the Stroop task – we found that the processing of dominance signals cannot be ignored. Importantly, this is in line with previous social-rank studies assessing related components of automaticity such as involuntary attention (eye-tracking, DeWall & Maner, 2008; Maner et al., 2008), automatic evaluation (a variant of the affective Simon paradigm, Moors & De Houwer, 2005) or elicitation of brain response (Zink et al., 2008). Second, we included vocal signals of dominance while most behavioural studies focused solely on the processing of visual information. Finally, previous findings have relied exclusively on unimodal stimuli despite the ubiquity of multimodal stimulation during social interactions. The inclusion of such ecologically valid stimuli enhances the relevance of our findings to everyday communication.

These findings are also in line with a growing number of studies suggesting that information from one modality interacts with information conveyed by another modality in affecting interpersonal perception. In addition to the already discussed case of emotion congruence, the congruence (vs. incongruence) in the familiarity of faces and voices improves person recognition (O'Mahony & Newell, 2012; Schweinberger, Robertson, & Kaufmann, 2007). Another example of cross-modal interference includes the processing of gender (Smith, Grabowecky, & Suzuki, 2007). For instance, Smith et al. (2007) reported that androgynous faces are more likely to be judged as masculine when combined with tones in the male fundamental-speaking-frequency range, whereas when these faces were presented with tones in the female fundamental-speaking-frequency range, they were more likely to be perceived as feminine. In sum, the processing of social-rank cues appears to resemble the rapid processing of other salient social characteristics, and suggest that social-rank is a significant feature extracted from visual and vocal information.

Moreover, our data suggest that ignoring vocal cues was easier than ignoring facial cues, as indicated by faster RTs and greater decision accuracy in the face-attend than voice-attend tasks. These data dovetail with previous findings showing better performance when participants attended to the emotional facial expressions compared to the vocal ones (Collignon et al., 2008). Interestingly, a study using a bimodal Stroop task with emotional faces and voices suggests that this effect may be moderated by culture (Liu, Rigoulot, & Pell, 2015b); the authors demonstrated that both adults from East Asian and Western cultures were sensitive to emotional incongruity between channels. However, whereas Chinese participants did not show significant differences between the face-attend and voice-attend tasks, English participants showed greater interference from the to-be-ignored irrelevant faces than by the to-be-ignored irrelevant voices. Our results from an Israeli sample thus align with the Western pattern reported by Liu, Rigoulot, and Pell's findings (see also, Liu, Rigoulot, & Pell, 2015a; Tanaka et al., 2010).

In closing, we highlight the limitations of the present study, and sketch some areas of future research. *First*, we focused solely on social-rank traits. Future studies may want to investigate trustworthiness, which has also been identified as a critical social dimension along which faces and voices are automatically evaluated (McAleer et al., 2014; Oosterhof & Todorov, 2008; Sutherland et al., 2013). *Second*, we included only male faces and voices. Forthcoming studies may examine the effects of stimulus gender. *Third*, to provide a more complete picture of the interaction between visual and vocal modalities in the perception of dominance, future studies may incorporate other visual dominance cues, such as body postures and height. *Fourth,* it is possible that the greater influence of facial cues was driven by the specifics of the present study's design, rather than by the intrinsic features of the face processing. Indeed, our instructions, emphasising the need to attend to faces (in case the participants would be tempted to just close their eyes and press the relevant buttons while listening to the voices) may have enhanced facial features processing. In addition, it is also possible that greater subjective difference in the social-rank ratings between the faces and the voices may be responsible for this effect. To address these possible confounds, future studies may conduct experiments with more balanced instructions for the face-attend and voice-attend tasks, as well as with more subjective consistency between the facial and the vocal stimuli. *Finally,* the role of individual differences in social dominance perception can profitably be examined. For example, social anxiety has been characterised by a hypersensitivity to social dominance (see Gilboa-Schechtman & Shachar-Lavie, 2013). Extending those studies to performance-based attentional measures may differentiate between sensitivity to, and explicit evaluation of, interpersonal dominance signals. Moreover, the examination of cross-modal stimuli may also inform the respective influence of different sensory channels on the processing of interpersonal information in social anxiety (Peschard, Maurage, & Philippot, 2014).

In conclusion, our results suggest that, like gender, emotion and familiarity, social-rank cues as conveyed by facial and vocal information are processed in a relatively automatic fashion. The centrality of the social-rank bio-behavioural system for the processing of ecologically salient interpersonal cues is therefore highlighted.

Acknowledgements

The authors would like to thank Hannah Restle and Hadar Keshet for their useful comments and suggestions on earlier drafts of this paper as well as Sivan Ben-Moshe, Omri Rubinstein and Saray Peled-Maderer for their help in collecting data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Israel Science Foundation [grant number 740/15] awarded to Eva Gilboa-Schechtman.

References

Cheng, J. T., & Tracy, L. J. (2014). Toward a unified science of hierarchy: Dominance and prestige are two fundamental pathways to human social rank. In J. T. Cheng, L. J. Tracy, & C. Anderson (Eds.), *The psychology of social status* (pp. 3–27). New York, NY: Springer.

Cheng, J. T., Tracy, J. L., Ho, S., & Henrich, J. (2016). Listen, follow me: Dynamic vocal signals of dominance predict emergent social rank in humans. *Journal of Experimental Psychology: General*, *145*(5), 536–547. doi:10.1037/xge0000166

Chiao, J. Y., Adams, R. J., Tse, P. U., Lowenthal, W. T., Richeson, J. A., & Ambady, N. (2008). Knowing who is boss: fMRI and ERP investigations of social dominance perception. *Group Processes & Intergroup Relations*, *11*(2), 201–214. doi:10.1177/1368430207088038

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, *1242*, 126–135. doi:10.1016/j.brainres.2008.04.023

Dalmaso, M., Galfano, G., Coricelli, C., & Castelli, L. (2014). Temporal dynamics underlying the modulation of social status on social attention. *Plos One*, *9*(3), e93139, 1–7. doi:10.1371/journal.pone.0093139

Dalmaso, M., Pavan, G., Castelli, L., & Galfano, G. (2012). Social status gates social attention in humans. *Biology Letters*, *8*(3), 450–452. doi:10.1098/rsbl.2011.0881

DeWall, C. N., & Maner, J. K. (2008). High status men (but not women) capture the eye of the beholder. *Evolutionary Psychology*, *6*(2), 328–341. doi:10.1177/147470490800600209

Feng, C., Tian, T., Feng, X., & Luo, Y. J. (2015). Attention modulations on the perception of social hierarchy at distinct temporal stages: An electrophysiological investigation. *Neuropsychologia*, *70*, 21–29. doi:10.1016/j.neuropsychologia.2015.02.013

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, *117*(3), 319–331. doi:10.1016/j.cognition.2010.09.003

Fournier, M. A., Moskowitz, D. S., & Zuroff, D. C. (2002). Social rank strategies in hierarchical relationships. *Journal of Personality and Social Psychology*, *83*(2), 425–433. doi:10.1037/0022-3514.83.2.425

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14*(3), 289–311. doi:10.1080/026999300378824

Gilboa-Schechtman, E., & Shachar-Lavie, I. (2013). More than a face: A unified theoretical perspective on nonverbal social cue processing in social anxiety. *Frontiers in Human Neuroscience*, *7*, 904, 1–14. doi:10.3389/fnhum.2013.00904

Johnson, S. L., Leedom, L. J., & Muhtadie, L. (2012). The dominance behavioral system and psychopathology: Evidence from self-report, observational, and biological studies. *Psychological Bulletin*, *138*(4), 692–743. doi:10.1037/a0027503

Jones, B. C., DeBruine, L. M., Main, J. C., Little, A. C., Welling, L. L., Feinberg, D. R., & Tiddeman, B. P. (2010). Facial cues of dominance modulate the short-term gaze-cuing effect in human observers. *Proceedings of the Royal Social: Biological Sciences*, *277*(1681), 617–624. doi:10.1098/rspb.2009.1575

Klofstad, C. A., Anderson, R. C., & Nowicki, S. (2015). Perceptions of competence, strength, and age influence voters to select leaders with lower-pitched voices. *PLos One*, *10*(8), e0133779, 1–14. doi:10.1371/journal.pone.0133779

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society: Biological Sciences*, *279*(1738), 2698–2704. doi:10.1098/rspb.2012.0311

Liu, P., Rigoulot, S., & Pell, M. D. (2015a). Cultural differences in on-line sensitivity to emotional voices: Comparing east and west. *Frontiers in Human Neuroscience*, *9*, 311, 1–12. doi:10.3389/fnhum.2015.00311

Liu, P., Rigoulot, S., & Pell, M. D. (2015b). Culture modulates the brain response to human expressions of emotion: Electrophysiological evidence. *Neuropsychologia*, *67*, 1–13. doi:10.1016/j.neuropsychologia.2014.11.034

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. doi:10.3758/s13428-014-0532-5

Main, J. C., Jones, B. C., DeBruine, L. M., & Little, A. C. (2009). Integrating gaze direction and sexual dimorphism of face shape when perceiving the dominance of others. *Perception*, *38*(9), 1275–1283. doi:10.1068/p6347

Maner, J. K., DeWall, C. N., & Gailliot, M. T. (2008). Selective attention to signs of success: Social dominance and early stage interpersonal perception. *Personality and Social Psychology Bulletin*, *34*(4), 488–501. doi:10.1177/0146167207311910

Mazur, A. (2005). *Biosociology of dominance and deference*. Lanham, MD: Rowman & Littlefield.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say "hello"? Personality impressions from brief novel voices. *PLos One*, *9*(3), e90779, 1–9. doi:10.1371/journal.pone.0090779

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. doi:10.1037/0033-2909.109.2.163

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review*, *110*(3), 422–471. doi:10.1037/0033-295X.110.3.422

Moors, A., & De Houwer, J. (2005). Automatic processing of dominance and submissiveness. *Experimental Psychology*, *52*(4), 296–302. doi:10.1027/1618-3169.52.4.296

Mueller, U., & Mazur, A. (1996). Facial dominance of west point cadets as a predictor of later military rank. *Social Forces*, *74*(3), 823–850. doi:10.1093/sf/74.3.823

O'Mahony, C., & Newell, F. N. (2012). Integration of faces and voices, but not faces and names, in person recognition. *British Journal of Psychology*, *103*(1), 73–82. doi:10.1111/j.2044-8295.2011.02044.x

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of National Academy of Sciences of the USA*, *105*(32), 11087–11092. doi:10.1073/pnas.0805664105

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, *35*(2), 192–201. doi:10.1007/s11031-011-9206-0

Peschard, V., Maurage, P., & Philippot, P. (2014). Towards a cross-modal perspective of emotional perception in social anxiety: Review and future directions. *Frontiers in Human Neuroscience*, *8*, 322, 1–6. doi:10.3389/fnhum.2014.00322

Pun, A., Birch, S. A., & Baron, A. S. (2016). Infants use relative numerical group size to infer social dominance. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(9), 2376–2381. doi:10.1073/pnas.1514879113

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296. doi:10.1016/j.evolhumbehav.2005.11.003

Ratcliff, N. J., Hugenberg, K., Shriver, E. R., & Bernstein, M. J. (2011). The allure of status: High-status targets are privileged in face processing and memory. *Personality and Social Psychology Bulletin*, *37*(8), 1003–1015. doi:10.1177/0146167211407210

Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior*, *39*(4), 355–370. doi:10.1007/s10919-015-0214-8

Santamaría-García, H., Burgaleta, M., & Sebastián-Gallés, N. (2015). Neuroanatomical markers of social hierarchy recognition in humans: A combined ERP/MRI study. *The Journal of Neuroscience*, *35*(30), 10843–10850. doi:10.1523/JNEUROSCI.1457-14.2015

Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *Quarterly Journal of Experimental Psychology (Hove)*, *60*(10), 1446–1456. doi:10.1080/17470210601063589

Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology*, *17*(19), 1680–1685. doi:10.1016/j.cub.2007.08.043

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. doi:10.1037/h0054651

Sussman, A. B., Petkova, K., & Todorov, A. (2013). Competence ratings in US predict presidential election outcomes in Bulgaria. *Journal of Experimental Social Psychology*, *49*(4), 771–775. doi:10.1016/j.jesp.2013.02.003

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. doi:10.1016/j.cognition.2012.12.001

Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I feel your voice: Cultural differences in the multisensory perception of emotion. *Psychological Science*, *21*(9), 1259–1262. doi:10.1177/0956797610380698

Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M., & Carey, S. (2011). Big and mighty: Preverbal infants mentally represent social dominance. *Science*, *331*(6016), 477–480. doi:10.1126/science.1199198

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. doi:10.1016/j.tics.2008.10.001

Vukovic, J., Jones, B. C., Feinberg, D. R., Debruine, L. M., Smith, F. G., Welling, L. L., & Little, A. C. (2011). Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women's preferences for masculine pitch in men's voices. *British Journal of Psychology*, *102*(1), 37–48. doi:10.1348/000712610X498750

Watanabe, N., & Yamamoto, M. (2015). Neural mechanisms of social dominance. *Frontiers in Neuroscience*, *9*, 154, 1–14. doi:10.3389/fnins.2015.00154

Wolff, S. E., & Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. *Behavioral Ecology and Sociobiology*, *64*, 1673–1683. doi:10.1007/s00265-010-0981-5

Zebrowitz, L. A. (1996). Physical appearance as a basis of stereotyping. In N. MacRae, M. Hewstone, & C. Stangor (Eds.), *Foundations of stereotypes and stereotyping* (pp. 79–120). New York, NY: Guilford Press.

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

Zebrowitz, L. A., & Montepare, J. M. (2015). Faces and first impressions. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Volume 1: Attitudes and social cognition* (pp. 251–276). Washington, DC: American Psychological Association.

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know your place: Neural processing of social hierarchy in humans. *Neuron*, *58*(2), 273–283. doi:10.1016/j.neuron.2008.01.025

Zitek, E. M., & Tiedens, L. Z. (2012). The fluency of social hierarchy: The ease with which hierarchical relationships are seen, remembered, learned, and liked. *Journal of Personality and Social Psychology*, *102*(1), 98–115. doi:10.1037/a0025345

Zuroff, D. C., Fournier, M. A., Patall, E. A., & Leybman, M. J. (2010). Steps toward an evolutionary personality psychology: Individual differences in the social rank domain. *Canadian Psychology*, *51*(1), 58–66. doi:10.1037/a0018472