The Multi-modal Nature of Trustworthiness Perception

Elena Tsankova [1], *Eva Krumhuber*[2], *Andrew J. Aubrey*[3], *Arvid Kappas*[1], *Guido Möllering*[1], *David Marshall*[3], *Paul L. Rosin*[3]

[1] Jacobs University Bremen, Germany
[2] University College London, UK
[3] Cardiff University, UK

etsankova@gmail.com, e.krumhuber@ucl.ac.uk, AubreyAJ@cardiff.ac.uk, a.kappas@jacobs-university.de, g.moellering@jacobs-university.de, Dave.Marshall@cs.cardiff.ac.uk, Paul.Rosin@cs.cf.ac.uk

Abstract

Most past work on trustworthiness perception has focused on the structural features of the human face. The present study investigates the interplay of dynamic information from two channels – the face and the voice. By systematically varying the level of trustworthiness in each channel, 49 participants were presented with either facial or vocal information, or the combination of both, and made explicit judgements with respect to trustworthiness, dominance, and emotional valence. For most measures results revealed a primacy effect of facial over vocal cues. In examining the exact nature of the trustworthiness - emotion link we further found that emotional valence functioned as a significant mediator in impressions of trustworthiness. The findings extend previous correlational evidence and provide important knowledge of how trustworthiness in its dynamic and multi-modal form is decoded by the human perceiver.

Index Terms: trustworthiness, face, voice, emotion, dynamic, multi-modal

1. Introduction

Trustworthiness is a highly desirable personality trait of significant evolutionary importance. Only if the other person is trustworthy are we confident in increasing our vulnerability by relying on them. In turn, if such quality is absent we may want to avoid possible costs of interaction due to fear of exploitation and deceit. Being able to accurately judge trustworthiness in others, therefore, constitutes an important skill for facilitating successful exchange and cooperation [e.g., 1]. Previous research has shown that humans are extremely efficient in detecting the relevant cues and evaluating the trustworthiness of strangers [e.g., 2].

Most of the trustworthiness research has been concerned extensively with facial cues and has disregarded other information channels such as the voice. In fact, the voice has recently been shown to be an important aspect of social signaling in trust research [e.g., 3]. But overall the knowledge that we have about the ways in which the voice influences perceived trustworthiness is rather limited in comparison to the information that exists about the face.

First impressions in real life, however, are not based solely on one information channel but rather rely on cues from multiple channels. This is especially important in the study of trust and deception, as it is likely that deceptive cues are conveyed via the channels over which we have little control [e.g., 4]. We may consequently miss important trustworthiness-related information if we are presented with only one information channel. Furthermore, a large body of evidence suggests that the accuracy of honesty and deception detection varies with the information channel. For example, the voice has been shown to be more important than the face in judging honesty [5]. Similarly, deception detection accuracy has been reported to be higher in the voice channel as compared to the body and face [6–7]. For the study of perceived trustworthiness it, therefore, seems to be essential to look at the two information channels together rather than in isolation, thereby allowing a more comprehensive view of the relative contribution of each.

To the best of our knowledge, however, there hasn't been a systematic investigation of trustworthiness impressions based on facial and vocal information at the same time. It is the purpose of the current research to investigate the formation of trustworthiness impressions in situations when the impression holder is presented simultaneously with both channels as opposed to only one information channel. By comparing the relative role of the facial and vocal component in the perception of trustworthiness, we aimed to test whether one information channel was more important than the other and whether trustworthiness impressions would benefit from combined in contrast to isolated channel presentation.

Up to now, the relative contribution of the two channels of interest (face and voice) to impression formation has been studied mainly in the context of the perception and recognition of emotions [e.g., 8–9], with findings speaking mainly in favor of the face as the dominating channel for emotional information transfer. Specifically, when sending conflicting information via the visual and auditory channel the face has been shown to receive disproportionately greater weight than the voice in determining the affective meaning of the message [e.g., 10] or attitude on part of the speaker [e.g., 11]. However, the issue of the relative channel contribution has not been specifically addressed so far in the context of trustworthiness information. This is surprising giving that the perception of trustworthiness is to a certain degree related to the perception of emotion [e.g., 12]. But the exact nature of the trustworthiness - emotion link is still unclear. Although it might fit descriptions of a correlation [13], there is the question of whether the inverse relationship between trustworthiness and emotion equally holds. Furthermore, how does the emotional valence (happy/angry) impact ratings of

trustworthiness, and can this process be described in terms of a mediational effect?

1.1. Aims of the present research

Our goal in this paper was two-fold. First, we set out to investigate the relative contribution of the two channels, face and voice, to the perception of trustworthiness. Second, we aimed to examine the exact nature of the relationship between trustworthiness and emotion reported by [13]. On the grounds of previous findings from the field of emotion research [e.g., 8–9] we predicted that facial information would primarily determine trustworthiness perceptions in the combined face & voice channel. At the same time, however, we expected participants to be more confident in their trustworthiness judgements when those were based on the vocal information or inconsistent cue combinations due to heavier reliance on the voice when the message is seen as deceptive [e.g., 5–7]. Furthermore, we predicted a differential contribution of the face and voice channels to perceived dominance, given previous findings showing a larger impact of the voice as compared to the face on perceived dominance [14].

For studying the relationship between trustworthiness and emotional valence, an approach similar to that by [12] was taken where perceivers were asked to make explicit trustworthiness judgements. However, instead of inferring emotional information from trustworthiness impressions based on neutral faces, we directly encoded trustworthiness in the stimuli and obtained ratings of perceived emotional valence. This allowed us to study the role emotion information more directly and to investigate whether emotional valence functions as a mediator in perceived trustworthiness.

2. Method

2.1. Participants

Forty-nine (36 female) White students at Jacobs University Bremen, age range 18 – 29 years (M = 20.40, SD = 1.98), took part in the experiment in return for a monetary reward (5 €) and, where applicable, partial course credit. Each participant was assigned to one of the three experimental conditions according to a predetermined sequence, resulting 16 participants in the Face & Voice condition, 16 participants in the Face only condition, and 17 participants in the Voice only condition.

2.2. Design

The experiment had a nested factorial design with one between-subject factor, Condition (Face, Voice, Face & Voice), and one within-subject factor, Channel Combination (trustworthy face + trustworthy voice or TF+TV, trustworthy face + non-trustworthy voice or TF+NTV, non-trustworthy face + trustworthy voice or NTF+TV, non-trustworthy face + non-trustworthy voice or NTF+NTV). Channel Combination was nested under Condition so that participants in the Face & Voice condition were presented with both channels in all four combinations, whereas participants in the Face and Voice groups were only presented with one channel for all combinations (facial channel in the Face only condition and vocal channel in the Voice only condition).

The four types of channel combination were obtained once for male and once for female targets, thus adding up to eight combinations. These combinations were shown a total of six times, each time with a different question measuring a dimension of interest. Therefore, in each condition there were a total of 48 stimuli, presented in a random order.

2.3. Stimuli

The stimuli were selected from a larger set [15] of short audiovisual clips of various human targets saying a standard sentence ("*Hello, my name is Jo*") in a trustworthy, non-trustworthy, and neutral manner. Channel transfer was achieved using a dynamic time warping technique [16], which involved the decoupling of the video and audio channels of the original recordings, followed by the systematic mapping of the trustworthy and non-trustworthy portrayals with the neutral recording. An example of the channel alignment is given in Figure 1. The algorithm guaranteed perfect synchronization of speech and lip movement to avoid confounding impressions of reduced realism. Based on the measured trustworthiness ratings of these aligned audio-visual stimuli in [15] we selected for the current experiment facial and auditory stimuli within each gender category that were judged as most and least trustworthy in each channel (face: M = 3.53 vs. M = 3.07, p < .01; voice: M = 4.07 vs. M = 2.74, p < .01, 7-point Likert scale). For these targets the video and audio channels were aligned on an inter-target basis within each gender category to obtain the following four combinations: TF+TV, TF+NTV, NTF+TV, and NTF+NTV. This resulted in a total of eight clips depicting consistent and inconsistent audio-visual trustworthiness combinations with an approximate duration of 3s each.

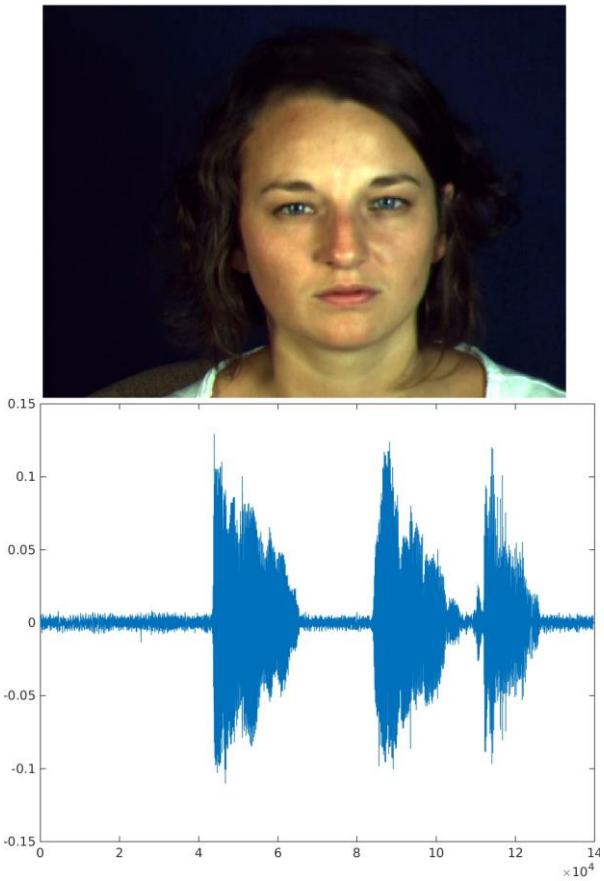


Figure 1: *Alignment of a non-trustworthy visual and a trustworthy auditory channel (NTF + TV).*

2.4. Procedure

Prior to their individual testing sessions all participants gave written informed consent. The experiment was computer-based. Stimuli in the Face only and Face & Voice conditions were presented on a black background, whereas in the Voice only condition participants solely listened to the targets' voices while looking at a blank black screen. For the clips including audio (Voice only, Face & Voice), sound was presented via Philips Stereo Headphones SBC HP090. Sound was muted in the Face only condition. All questions and response scales appeared on the screen once the clip had finished playing. After the experiment was competed participants were given the opportunity to ask questions related to the procedure and hypotheses. No participants reported noticing artifacts due to the channel alignment in the Face & Voice condition.

2.5. Dependent measures

The questions assessed three dimensions of interest: a) trustworthiness: "*How trustworthy is this person?*" and "*How insincere is this person?*", b) dominance: "*How dominant is this person?*"and "*How timid is this person?*", and c) emotional valence: "*How happy is this person?*" and "*How angry is this person?*". Presentation order of the questions was randomized with the exception that the question on perceived trustworthiness was always followed by participants' self-estimated confidence in the trustworthiness judgement ("*How confident are you in this answer?*"). Answers to all questions were obtained on seven-point Likert scales ranging from 1- *not at all* to 7 - *very much*.

3. Results

3.1. Analyses of variance

Due to the nested design and the between-subject manipulation of channel information, the data were analyzed separately for each condition. For the combined Face & Voice condition, a multivariate analysis of variance (MANOVA) with Channel Combination (TF+TV, TF+NTV, NTF+TV, NTF+NTV) as within-subjects factor was conducted on the dependent variables[1]. Because there was only one channel present in the Face only and Voice only condition, separate MANOVAs with Target Trustworthiness (T vs. NT) as a within-subject factor were performed.[2]

Significant multivariative effects emerged for all three conditions: Face & Voice, $F(7, 9) = 22.29$, $p < .001$; Face only, $F(7, 9) = 22.29$, $p < .001$; and Voice only, $F(7, 10) = 9.79$, $p = .001$. For all univariate analyses, a Greenhouse-Geisser adjustment to degrees of freedom was applied. The relevant means and standard errors for each information channel and condition are displayed in Figure 2.

On the univariate level, there was a significant effect of the *trustworthiness* measure for all three conditions: Face & Voice, $F(2.37, 35.49) = 19.41$, $p < .001$; Face, $F(1, 15) = 89.02$, $p < .001$; and Voice, $F(1, 16) = 7.63$, $p = .014$. A significant difference between the trustworthy and non-trustworthy stimuli emerged in the Face only and Voice only conditions, confirming that the manipulation of encoded trustworthiness was successful for each channel. When both channels were combined (Face & Voice condition) the facial component had a considerable stronger effect than the vocal component. Specifically, TF+TV and TF+NTV were judged as more trustworthy than NTF+TV (ps < .001) and NTF+NTV (ps < .001), suggesting that the relative impact of the vocal channel was discounted. Overall, ratings based on the combined channels closely mirrored those in the Face only condition. As predicted, self-estimated confidence in perceived trustworthiness differed significantly in the Voice only condition, with participants being more confident when rating trustworthy ($M = 5.10$, $SE = 0.34$) than non-trustworthy ($M = 4.46$, $SE = 0.30$) voices. But, no other effects emerged for ratings of confidence.

In the case of the *insincerity* measure there was a significant effect in the Face & Voice condition, $F(2.46, 36.94) = 7.82$, $p = .001$, as well as in the two separate channel conditions: Face, $F(1, 15) = 12.87$, $p = .003$; and Voice, $F(1, 16) = 16.08$, $p = .001$. Post-hoc comparisons revealed results opposite to those for trustworthiness, with higher ratings for non-trustworthy compared to trustworthy stimuli in the Face only and Voice only condition. Again, the facial component was found to gain greater weight in the combined Face & Voice condition. Specifically, NTF+TV and NTF+NTV were seen as more insincere than TF+NTV (ps < .05) and TF+TV (ps < .01). However, the voice also emerged as influential component in the assessment of the total message, with more insincerity being attributed to TF+NTV than TF+TV (p = .045).

There was no significant variation as a function of the type of condition (Face only, Voice only) for ratings of *timidity*, Fs < 1.00, ps > 1.00. In the case of the Face & Voice condition there was only a tendency for an effect of timidity, $F(2.58, 38.66) = 2.50$, $p = .082$. Similarly, in the Voice only condition, trustworthy voices were not rated differently from non-trustworthy voices with respect to perceived *dominance* ($F(1, 16) = 1.92$, $p = .185$). There was, however, an effect of dominance in the Face & Voice condition, $F(2.40, 36.03) = 12.78$, $p < .001$, and the Face only condition, $F(1, 15) = 23.71$, $p < .001$. Perceptions of dominance were higher for non-trustworthy compared to trustworthy faces in the Face only condition. When both channels were combined (Face & Voice), there was a steady increase in dominance ratings. Contrary to predictions, the facial component formed the primacy basis for evaluations, with TF+TV and TF+NTV being judged as less dominant than NTF+NTV (ps < .001), and TF+TV receiving lower ratings than NTF+TV ($p = .004$). Nevertheless, the vocal component led to a marginal significant discrimination between NTF+TV and NTF+NTV ($p = .053$) with most dominance being attributed to the latter channel combination.

With respect to perceived *happiness*, an effect was present in all three conditions: Face & Voice, $F(1.99, 29.83) = 21.31$, $p < .001$; Face, $F(1, 15) = 82.28$, $p < .001$; and Voice: $F(1, 16) = 56.07$, $p < .001$. Post-hoc comparisons showed that trustworthy stimuli attracted significantly higher ratings than non-trustworthy stimuli in the Face only and Voice only condition. In the decoding of the combined message (Face & Voice), ratings were again closer to the Face only compared to the Voice only condition. That is, participants based their attributions regarding the impression target's affective state predominantly on facial cues. Specifically, TF+TV and TF+NTV were judged as happier than NTF+TV (ps < .01) and NTF+NTV (ps < .001). Furthermore, there was a significant difference between NTF+TV and NTF+NTV ($p = .038$), pointing to some influence of the vocal component.

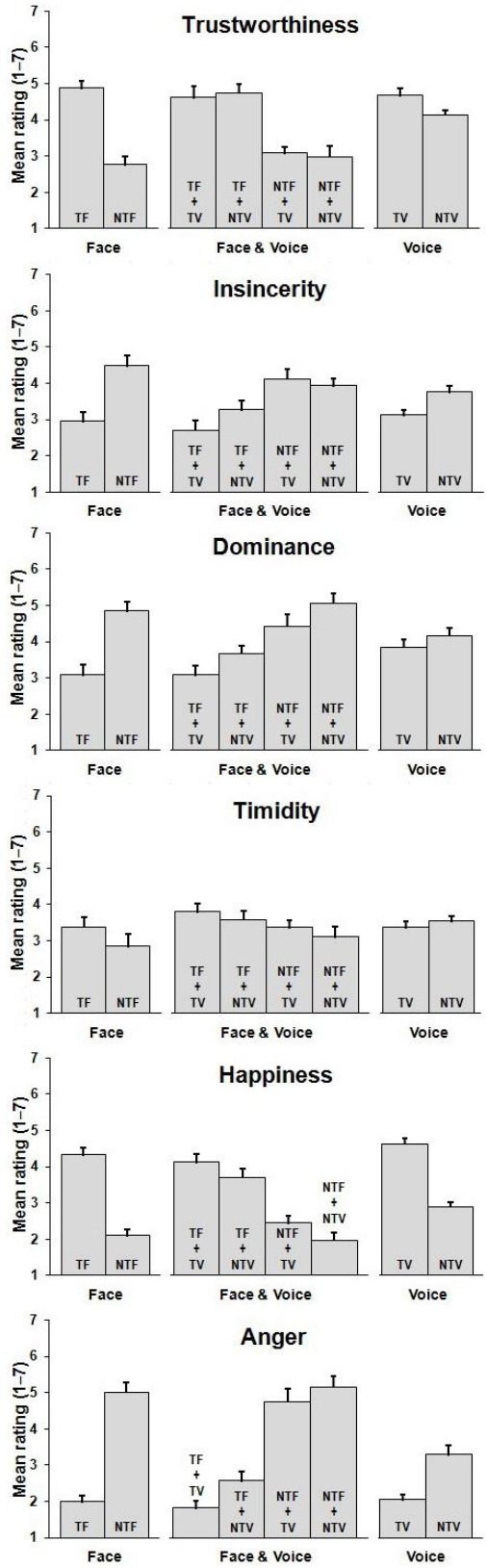


Figure 2: *Mean ratings for six dependent measures in the single (Face, Voice) and combined (Face &Voice) information channels. Error bars indicate + 1 standard error of the mean. TF = trustworthy face, TV = trustworthy voice, NTF = non-trustworthy face, NTV = non-trustworthy voice.*

The results were highly similar, but opposite for attributions of *anger*. On the univariate level an effect of this measure was found in all three conditions: Face & Voice, $F(1.66, 24.85) = 36.46$, $p < .001$; Face, $F(1, 15) = 117.60$, $p < .001$; and Voice: $F(1, 16) = 24.15$, $p < .001$. Trustworthy stimuli were rated as less angry than non-trustworthy stimuli in the Face only and Voice only condition. When making judgements of the combined Face & Voice condition, the effect of the vocal component was reduced: Both NTF+TV and NTF+NTV were perceived as angrier than TF+NTV (ps < .01) and TF+TV (ps < .001). Although the facial channel received greater weight in the decoding of the combined message, the vocal component was found to contribute to the discrimination between TF+TV and TF+NTV ($p = .013$).

On the whole, the effects of the facial and vocal component in the Face & Voice condition seemed to follow a linear trend (TF+TV, TF+NTV, NTF+TV, NTF+NTV). To test whether the data fit a linear line, polynomial contrast analyses were conducted for the dimensions of interest. Results revealed significant linear trends for trustworthiness, $F(1, 15) = 27.76$, $p < .001$, insincerity, $F(1, 15) = 13.22$, $p = .002$, dominance, $F(1, 15) = 35.17$, $p < .001$, timidity, $F(1, 15) = 6.34$, $p = .024$, happiness, $F(1, 15) = 34.78$, $p < .001$, and anger, $F(1, 15) = 82.67$, $p < .001$, confirming that judgement ratings varied linearly with the channel combinations.[3]

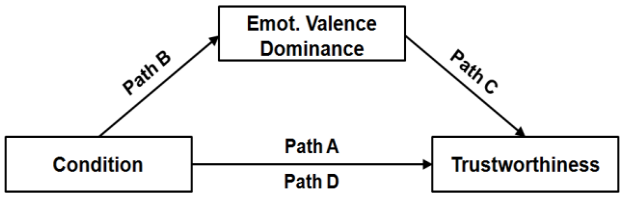
3.2. Channel weighting

Regression analyses examined the relative contribution of each single channel (Face only, Voice only) to ratings in the combined Face & Voice channel. Results showed that the Face only channel accounted for approximately 9 times as much variance as the Voice only channel. Overall, the variances explained were 27.2% by face and 1.6% by voice for trustworthiness, 7.4% by face and 0.1% by voice for insincerity, 12.9% by face and 5.8% by voice for dominance, 1.5% by face and 1.6% by voice for timidity, 26.8% by face and 3.7% by voice for happiness, and 57.9% by face and 2% by voice for anger. In line with previous research, variability of responses to stimuli in the Face & Voice condition consequently seemed to be determined primarily by variations in the face.

3.3. Mediational analyses

To test the prediction that differences in trustworthiness impressions of the three conditions would be mediated by the perceived emotional valence of the stimuli, mediation analyses were conducted following [17]. As summarized in Table 1, stimulus condition significantly predicted ratings of trustworthiness, as well as of emotional valence (happy/angry). Similarly, emotional valence was a significant predictor of how trustworthy a person was judged to be. However, when controlling for the effect of emotional valence, stimulus condition no longer predicted ratings of trustworthiness. Sobel's test [18] was significant both for happiness and anger, showing that emotional valence successfully mediated perceived trustworthiness. Although dominance was a significant predictor of trustworthiness ratings, the regression path from condition to perceived dominance was non-significant, thereby excluding its possibility for mediation. The emotional valence of a stimulus therefore appears to be the main determining factor in accounting for the perceived level of trustworthiness as a function of condition.

Table 1. Regression paths and coefficients including Sobel Test statistics for mediation analyses on perceived trustworthiness.



Mediator	*Path A*	*Path B*	*Path C*	*Path D*	*z*
Emot. valence					
Happiness	.28** (.10)	.27* (.11)	.61*** (.05)	.13 (.08)	2.39
Anger	.28** (.10)	.43** (.14)	.47*** (.04)	.09 (.08)	2.96
Dominance	.28** (.10)	.01 (.11)	.37*** (.06)	.29* (.09)	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Values in parentheses indicate the standard error of the unstandardized regression coefficient.

4. Discussion

In this study we investigated the relative contribution of two information channels, face and voice, in explicit judgements of trustworthiness and related trait characteristics. By using a dedicated experimental design that featured consistent and inconsistent combinations of facial-vocal communication we showed that trustworthiness impressions were more influenced by the facial as compared to the vocal component. These findings are in line with previous research demonstrating a stronger effect of the face for attributions of emotions [e.g., 8–9]. Trustworthiness information therefore seems to be distributed in a similar manner as emotional information among the two perceptual modalities. Moreover, such video primacy also applied for ratings of dominance, contrary to our expectations based on [14]. Given that there exists a strong (negative) association between trustworthiness and aggression in face perception [12–13], we suspect that dominance may have functioned in this study as a proxy for aggression related traits, with more weight being assigned to the same channel as for rating trustworthiness. Such explanation is also supported by the greater reliance on the face found for judgements of anger, resembling aggression and dominance [19].

Despite the primary role of the face, the effects due to the vocal component were not entirely redundant. In fact, participants were more confident when rating trustworthy as opposed to non-trustworthy vocal messages, thereby providing supportive evidence that perceivers did sense potential deception from the voice [5–7]. When examining the relative contribution of the face and voice, however, ratings from the combined Face & Voice channel were a linear function of the ratings in each component, with the face channel receiving 10/9 times the weight received by the vocal channel. In mixed channel communications, the decoding of trustworthy/non-trustworthy stimuli therefore appears to be most similar to that when being exposed to the face only.

Encoded trustworthiness in both channels was linked to the perception of emotion. In line with [13] trustworthy faces were perceived as happier than non-trustworthy faces and non-trustworthy faces were perceived as angrier than trustworthy faces. This finding was also true for the perception of the voice channel – trustworthy voices were perceived as happier than non-trustworthy voices and non-trustworthy voices were perceived as angrier than trustworthy voices. This pattern of results could be explained by the fact that dynamic facial properties carry information about motivation [e.g., 20], suggesting the intentions of the target – someone appearing or sounding happy is more likely to have positive intentions and act in a trustworthy manner than someone appearing and sounding angry. By studying the relationship between trustworthiness and emotion we demonstrated in this study that emotional valence (but not dominance) functioned as a mediator in trustworthiness impressions. This goes beyond existing evidence that although well-validated has been merely correlational in nature [e.g., 13] and suggests a major auxiliary role of emotion and its expression in judgements of trustworthiness and interpersonal trust [see 1]. Furthermore, it extends findings of a shared perceptual basis of emotional valence and trustworthiness [12–13] by showing respective effects with human-realistic stimuli displaying life-like behaviour rather than using computer-generated faces.

While strong effects of structural features on rapid impressions of trustworthiness could be previously shown, much less is known when multimodal dynamic information is available. The present study is the first to provide insights into how dynamic concordant and discordant multimodal information affects the attribution of trustworthiness and related person characteristics. Related to the dynamic nature of the information are two limitations of the present study. First, the present design did now allow for differentiation between the roles of facial shape and motion in the resulting trustworthiness impression. Thus, it still remains unclear what it is precisely within the facial channel that constitutes such a powerful cue for trustworthiness perception. Second, it is likely that verbal content contributes to the resulting trustworthiness impression [e.g., 8]. However, in the current study verbal content was kept constant and void of trustworthiness-related information, which could partially explain the observed low reliance on the vocal channel in the perception of trustworthiness. Future research on the multi-modal nature of trustworthiness would, therefore, benefit from addressing the relative roles of verbal, facial structural, and facial dynamic information.

In summary, the current study provided a first attempt to investigate the multi-modal nature of trustworthiness. By systematically varying the relevant information in the facial and vocal channel, we showed that the face received considerably greater weight in perceptions of trustworthiness and such impressions were mediated by the perceived emotional valence of the stimulus. These findings offer new perspectives for the understanding of the complex dynamic interplay among perceptual modalities and in turn contribute to our knowledge of how trustworthiness in its dynamic and multi-modal form is decoded by the human perceiver.

5. Acknowledgements

This research was conducted with the support of the European Commission under grant agreement eCute (FP7-ICT-2009-5) as well as the research center CoWell at Jacobs University Bremen.

6. References

[1] R. Boone and R. Buck, “Emotional expressivity and trustworthiness: The role of nonverbal behaviour in the

evolution of cooperation", *J. Nonverbal Behav.*, vol. 27, pp. 163–182, 2003.
[2] A. Todorov and J. Willis, "First impressions: Making up your mind after a 100-ms exposure to a face", *Psychol. Sci.*, vol. 17, no. 7, pp. 592–598, 2006.
[3] B. Waber, M. Williams, J. S. Carroll, and A. S. Pentland, "A voice is worth a thousand words: the implications of the micro-coding of social signals in speech for trust research", in *Handbook of Research Methods on Trust*, F. Lyon, G. Möllering, and M. N. K. Saunders, Eds. Cheltenham: Edward Elgar Publishing Limited, 2012, pp. 249–258.
[4] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiat.*, vol. 32, pp. 88–106, 1969.
[5] M. Zuckerman, R. DeFrank, J. Hall, D. Larrance, and R. Rosenthal, "Facial and vocal cues of deception and honesty", *J. Exp. Soc. Psychol.*, vol. 15, pp. 378–396, 1979.
[6] M. Zuckerman, B. M. DePaulo, and R. Rosenthal, "Verbal and nonverbal communication in deception", in *Advances in Experimental Social Psychology*, vol. 14, L. Berkowitz, Ed. New York: Academic, 1981, pp. 1–60.
[7] A. Manstead, H. Wagner, and C. MacDonald, "Face, Body, and Speech as Channels of Communication in the Detection of Deception", *Basic Appl. Soc. Psych*., vol. 5, pp. 317–332, 1984.
[8] U. Hess, A. Kappas, and K. R. Scherer. "Multichannel communications of emotion: Synthetic signal production," in *Facets of Emotion: Recent Research*. K. R. Scherer, Ed. Hillsdale, NJ: Erlbaum, 1988, pp. 161–182.
[9] E. Krumhuber and A. Manstead, "Are you joking? The moderating role of smiles in the perception of verbal statements", *Cogn. Emot.*, vol. 23, pp. 1504–1515, 2009.
[10] M. Argyle, F. Alkema, and R. Gilmour, "The communication of friendly and hostile attitudes by verbal and non-verbal signals," *Eur. J. Soc. Psychol.*, vol. 1, pp. 385–402, 1971.
[11] A. Mehrabian and S. Ferris, "Inference of attitudes from nonverbal communication in two channels", *J. Consult. Psych.*, vol. 31, no. 3, pp. 248–252, 1967.
[12] N. Oosterhof and A. Todorov, "The functional basis of face evaluation", *P. Natl. Acad. Sci.*, vol. 105, pp. 11087–11092, 2008.
[13] N. Oosterhof and A. Todorov, "Shared perceptual basis of emotional expressions and trustworthiness impressions from faces", *Emotion*, vol. 9, pp. 128–133, 2009.
[14] M. Zuckerman and R. Driver, "What sounds beautiful is good: The vocal attractiveness stereotype", *J. Nonverbal Behav*., vol. 13, pp. 67–82, 1988.
[15] E. Tsankova, A. J. Aubrey, E. Krumhuber, G. Mollering, A. Kappas, D. Marshall, and P. L. Rosin, "Facial and vocal cues in perception of trustworthiness", in *Lecture Notes in Computer Science: Vol. 7729. Computer Vision – ACCV 2012 Workshops Part II*, Berlin and Heidelberg, 2013, pp. 308–319.
[16] Aubrey, V. Kajic, I. Cingovska, P. Rosin, and D. Marsall, "Mapping and manipulating facial dynamics", in *Proceedings of the Institute of Electrical and Electronics Engineers Conference on Automatic Face and Gesture Recognition*, 2011, pp. 639–645.
[17] R. M. Baron and D. A Kenny, "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations", *J. Pers. Soc. Psychol.*, vol. 51, pp. 1173–1182, 1986.
[18] M. E. Sobel. "Asymptotic intervals for indirect effects in structural equation models", in *Sociological Methodology*, S. Leinhart, Ed. San Francisco: Jossey–Bass, 1982, pp. 290–312.
[19] J. M. Monterpare and H. Dobish, "The contribution of emotion perceptions and their overgeneralizations to trait impressions", *J. Nonverbal Behav.*, vol. 27, pp. 237–254, 2003.
[20] A. Kappas, E. Krumhuber, and D. Küster, "Nonverbal communication", in *Facial Behavior*, J. A. Hall and M. L. Knapp, Eds. Berlin & New York: de Gruyter/Mouton, 2013, pp. 131–166.

7. Footnotes

[1] Data reduction was attempted by combining the two items of each dimension (reverse scoring insincerity, timidity, anger) into one scale (trustworthiness, dominance, emotional valence). However, due to insufficient scale consistency all seven items are treated separately in the analyses.

[2] A single MANOVA with Condition as between-subjects factor was considered but rejected due to the presence of only one information channel in the Face only and Voice only condition, thus making it impossible to directly compare the channel combinations for all three conditions as some combinations were simply missing in the Face only and Voice only conditions. .

[3] In addition, a cubic trend also emerged for trustworthiness, $F(1, 15) = 18.85$, $p = .001$, and anger, $F(1, 15) = 4.90$, $p = .043$. The latter could be explained by the fact that the increase in trustworthiness and anger was most pronounced when the facial component of combined messages (Face & Voice) was trustworthy or non-trustworthy, respectively.