

Facial and Vocal Cues in Perceptions of Trustworthiness

Elena Tsankova¹, Andrew J. Aubrey², Eva Krumhuber¹, Guido Mllering¹,
Arvid Kappas¹, David Marshall², Paul L. Rosin²

¹ School of Humanities and Social Sciences, Jacobs University Bremen,
Campus Ring 1, 28759 Bremen, Germany

² School of Computer Science and Informatics, Cardiff University, Queens Buildings,
5 The Parade, Cardiff, CF24 3AA, United Kingdom

Abstract. The goal of the present research was to study the relative role of facial and acoustic cues in the formation of trustworthiness impressions. Furthermore, we investigated the relationship between perceived trustworthiness and perceivers confidence in their judgments. 25 young adults watched a number of short clips in which the video and audio channel were digitally aligned to form five different combinations of actors face and voice trustworthiness levels (neutral face + neutral voice, neutral face + trustworthy voice, neutral face + non-trustworthy voice, trustworthy face + neutral voice, and non-trustworthy face + neutral voice). Participants provided subjective ratings of the trustworthiness of the actor in each video, and indicated their level of confidence in each of those ratings. Results revealed a main effect of face-voice channel combination on trustworthiness ratings, and no significant effect of channel combination on confidence ratings. We conclude that there is a clear superiority effect of facial over acoustic cues in the formation of trustworthiness impressions, propose a method for future investigation of the judgmentconfidence link, and outline the practical implications of the experiment.

1 Introduction

Due to its evolutionary link to cooperation and survival, trust is a central aspect of human communication. Trust is typically established in the process of (continuous) interaction, but there is evidence that certain nonverbal information associated with trustworthiness can be processed already upon our very first encounter with an individual. For example, recent research by [1, 2] has shown that humans need no more than 100 ms to process facial features related to trustworthiness and generate an initial trustworthiness judgment. Such first impressions of trustworthiness, which emerge without the process of interpersonal interaction, are the focus of the current study. Here we investigate perceived trustworthiness, i.e., the subjectively estimated degree of another persons trustworthiness, rather than the actual trusting behavior in interaction. We also study how trustworthiness ratings are related to the persons own confidence in assessing the others trustworthiness.

In the above-mentioned research by Todorov and colleagues, perceived trustworthiness has been studied solely in the context of facial information. However, it has already been suggested that in addition to the visual channel, trustworthiness cues also travel via the auditory one [3]. The present study follows up on this and investigates the relative role of facial and vocal (paralinguistic, non-verbal) information in processing trustworthiness information. The study also looks at the way in which trustworthiness cues from one information channel (visual or auditory) bias perception of the information presented in the other channel. There is no prior empirical evidence on whether visual cues are more important than auditory cues for perceived trustworthiness, how they interact, and how people deal with inconsistent cues on different channels.

Not least because visual and auditory cues can be inconsistent, people may be able to form trustworthiness judgments quickly, but their confidence in those judgments may vary. Our study enables us to test whether consistent cues are associated with higher confidence ratings. More interestingly, though, by including confidence in ones assessment as a construct besides the perceived trustworthiness of the other, we can connect this psychological experiment to the more sociological-philosophical literature that emphasizes the leap of faith as an element in producing a state of trust [4–7]. Essentially, this literature claims that trust goes beyond the available information [8, 9]. In the context of our study this could mean that truly trusting individuals (i.e., those suspending uncertainty) generally give higher trustworthiness ratings, even when they are not fully confident about their assessment of the other (e.g., due to cue inconsistency). Alternatively, it could mean that highly confident individuals generally give higher trustworthiness ratings, even when they receive inconsistent cues. Both possibilities imply that the ‘leap of faith’ does not come as a result of the cognitive processing of trustworthiness information but is already part of this process. Our study explores whether the ‘leap of faith’ phenomenon can indeed be empirically observed and whether it is associated with certain characteristics of the visual and auditory cues or with certain types of study participants.

2 Predictions

The predictions we put forward in this study were twofold: first regarding the relationship between perceived trustworthiness and different combinations of visual and auditory cues; second about the relationship between confidence and perceived trustworthiness, and possible ‘leap-of-faith’ evidence. With respect to the relative role of facial and acoustic information in trustworthiness judgments, we took an explorative approach. We already knew from [10] that trustworthiness information is conveyed via facial dynamics. However, it is not so trivial to put forward a specific hypothesis about the degree to which visual and auditory information affect the perception of trustworthiness. There exists a large body of research on cross-modal integration in person identification ([11, 12] and many others), but not so much on face-voice information interaction in the case of person evaluation. But since trustworthiness has been shown to be

closely related to emotional valence [13], we expected that previous findings on cross-modal integration in the case of emotion expression would apply to the perception of trustworthiness as well. [14] showed that both facial and vocal information play a role in the communication of emotion, but the influence of facial information is bigger. Based on their report one would expect that in the present study facial information will bias trustworthiness judgments in the following way: The trustworthy face + neutral voice combinations would receive the highest trustworthiness ratings, followed by the remaining combinations in the order neutral face + trustworthy voice, neutral face + neutral voice, neutral face + non-trustworthy voice, and non-trustworthy face + neutral voice.

With respect to the relationship between confidence and perceived trustworthiness, we did not expect a linear correlation but rather a bifurcation effect whereby individuals would generally give more extreme trustworthiness ratings (very high or very low) when they are confident in their judgments and would tend towards more moderate ratings when they are not confident in their judgments. It should be noted at this point that we expect overall lower confidence in cases where the visual and acoustic trustworthiness cues are inconsistent. For the ‘leap of faith’ argument, we would test whether high trustworthiness ratings occur in spite of lower confidence ratings or in spite of inconsistent cues and under which conditions this happens, e.g. specific patterns in terms of various cue combinations. We were not aware of any experimental study that has analyzed these effects.

3 Method

3.1 3.1 Participants

Data were collected from 25 (14 female) representatives of the student population of Jacobs University Bremen, age range 18 – 27 years ($M = 19.76$, $SD = 2.03$), who volunteered to take part in this study in return for monetary compensation (5€ per participant) and optional partial credit for one of two introductory methods courses. Informed consent was obtained from all participants prior to their respective individual testing sessions.

3.2 Stimuli

The stimulus material used to generate trustworthiness evaluations consisted of short audiovisual clips of various encoders saying the sentence “Hello, my name is Jo”. This sentence was chosen because it is gender-neutral and does not contain any inherent trustworthiness cues. Thus, verbal content can be ignored and we can be relatively certain that trustworthiness cues in the audio channel are delivered via an acoustic route. Also, as it is an introductory phrase, this sentence adds a certain level of ecological validity to a situation of a first encounter with an unfamiliar person. Three types of recordings were obtained from 12 (six female) adult white Caucasian encoders. In one version encoders were instructed

to sound “as neutral as possible”, in the second version encoders were asked to sound trustworthy or “as if you are trying to convince somebody or if you want to make somebody believe you”, and in the third they were asked to sound non-trustworthy or “as if you are trying to deceive somebody”. It is important to note that none of the encoders came from the same student population as the target participant group. This measure was taken to prevent possible confounding effects of familiarity with the encoders as this could have potentially influenced trustworthiness evaluations. Female encoders wore no or minimal amounts of makeup and jewelry. There was variance with respect to facial hair in the case of male encoders, but that was considered a desirable phenomenon as it will allow us to study the effects of the presence or absence of facial hair on perceived trustworthiness.

The video and audio channels of each encoders recordings were then digitally aligned (audio channel was warped to match video channel where necessary) so that, in addition to the non-warped neutral face + neutral voice, the following four combinations were obtained: neutral face + trustworthy voice, neutral face + non-trustworthy voice, trustworthy face + neutral voice, and non-trustworthy face + neutral voice. These combinations were digitally warped [15] to achieve perfect synchronization of lip movement and voice. The neutral face + neutral voice recording was included as it was available from the neutral recording version, without alignment or warping. Thus, we ended up with a total of 60 stimulus clips (five for each of the 12 encoders). Roughly, the duration of the clips was around two seconds each. Stimuli in this study were presented on a computer via MediaLab [16] using Philips Stereo Headphones SBC HP090.

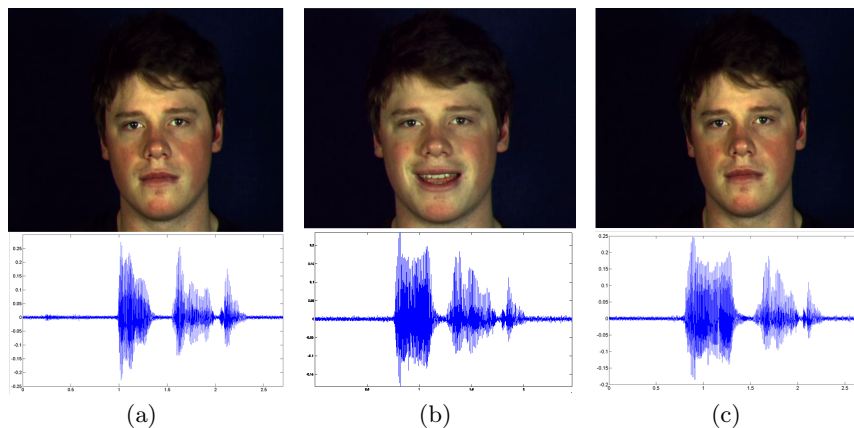


Fig. 1. Example of aligning (a) original neutral video and audio, with (b) original trustworthy video and audio. (c) neutral video with trustworthy audio. The trustworthy audio has been temporally aligned to the neutral video.

3.3 Procedure

Upon arrival at the lab participants were greeted by the experimenter and seated in the testing room, where they were asked to read and sign the informed consent form. The experimenter then gave a brief introduction to the study and left the room. Detailed instructions about the task were provided on the computer screen. Before the study began participants completed a short practice block to get used to the task (the encoder in the practice clips was different from the ones in the actual study, to avoid participants becoming familiar with one of the target persons as this could have an impact on trustworthiness ratings later on). The experimental task required participants to watch each clip carefully, then evaluate the trustworthiness of the person in the clip, and finally provide a confidence rating of their trustworthiness evaluation. The exact phrasing of the trustworthiness and confidence questions was respectively *How trustworthy is this person?* and *How confident are you in this answer?*. Responses were always given sequentially on a seven-point Likert scale ranging from *not at all* to *very much*. The videos were shown in a random order on a within-subject basis. After having seen all videos, participants were asked whether they recognized any of the encoders. Participants were not able to correctly identify any of the actors and it was, therefore, not necessary to exclude any data due to familiarity effects between encoders and decoders.

At the end of the study, basic demographic information about age, gender, country of origin, and ethnicity was obtained from each participant. Upon completion of the experimental session participants were thanked and reimbursed for their help, and were given the opportunity to ask questions regarding the study.

4 Results

All ratings on the variables of interest were converted to standardized scores. Three participants who scored beyond the range ± 2.5 were treated as outliers and were not included in any further analyses reported here. The current report is, therefore, based on the data of 22 participants (12 female), age range 18–27 years ($M = 19.82$, $SD = 2.15$).

To address the questions of interest we computed a trustworthiness and a confidence score for all encoders taken together. The trustworthiness score was the mean average trustworthiness rating across all actors, and the confidence score was the mean average confidence rating. Both trustworthiness and confidence scores were obtained for all five video-audio stimulus combinations.

In the next step of the analyses we conducted a repeated measures analysis of variance (RM-ANOVA) on the trustworthiness scores with video-audio combination entered as a within-subject factor. The results indicated a main effect of channel combination on the mean trustworthiness ratings, $F(4, 18) = 27.70$, $p < .001$, $\eta_p^2 = .86$ (see Figure 2). Pairwise comparisons showed that the trustworthy face + neutral voice condition was judged as being most trustworthy, with

ratings significantly different from the other conditions ($ps < .01$). The conditions in which the trustworthiness of the voice was varied (e.g., neutral face + trustworthy voice and neutral face + non-trustworthy voice) did not differ significantly from each other and from the neutral face + neutral voice condition ($ps > .05$). These three conditions (neutral face + trustworthy voice, neutral face + non-trustworthy voice, and neutral face + neutral voice) all received higher trustworthiness ratings than the non-trustworthy face + neutral voice condition ($ps < .01$).

We further computed a RM-ANOVA on the confidence scores (again, with video-audio combination as a within-subject factor), but this analysis revealed no significant effect of channel combination on participants confidence ratings, $p > .1$.

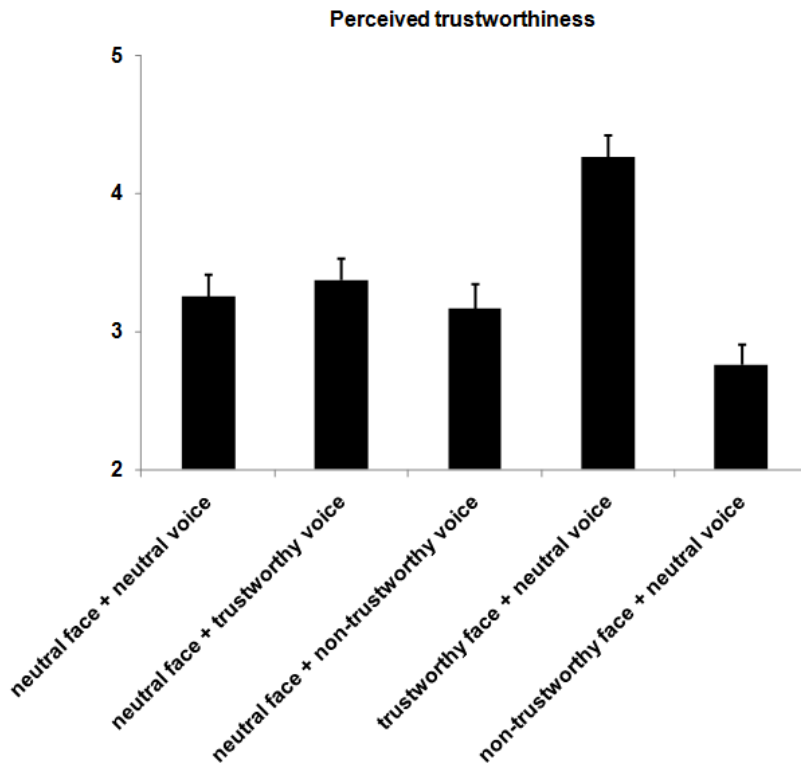


Fig. 2. Mean average trustworthiness ratings per channel combination. Error bars indicate 1 Standard Error.

With view to the ‘leap-of-faith’ hypothesis we plotted the trustworthiness scores against the confidence scores for each channel combination (see Figure 3).

In this way we could obtain a more comprehensive view of our data distribution. Clearly, participants tended towards high confidence ratings and low to medium perceived trustworthiness ratings in all channel combinations. This data pattern posed a challenge to testing the predicted bifurcation model since there were no data points reflecting extremely low confidence or extremely high trustworthiness ratings. It did, however, seem likely that the relationship between perceived trustworthiness and confidence scores might follow a linear trend. We tested this possibility in a simple linear regression model, where confidence was entered as an independent variable and perceived trustworthiness as a dependent variable. The test was performed for all five channel combinations, but returned a significant result only in the case of the non-trustworthy face + neutral voice condition, $b = .42, t(20) = 2.07, p = .05$. In that case confidence score explained some of the variance in perceived trustworthiness, $R^2 = .18, F(1, 20) = 4.28, p = .05$.

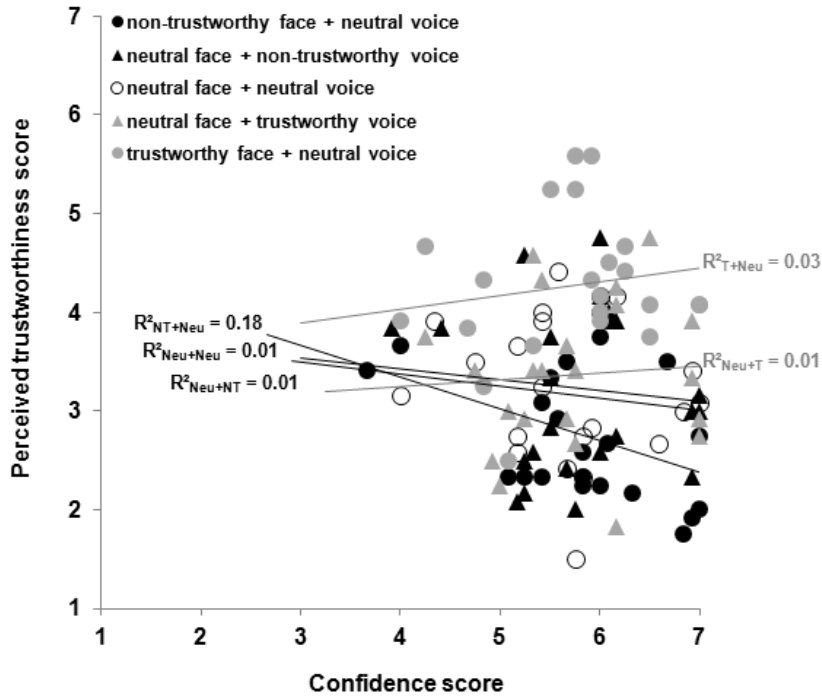


Fig. 3. Confidence scores plotted against perceived trustworthiness scores for all channel combinations.

5 Discussion

With respect to the relative role of visual and auditory cues in explicit judgments of trustworthiness, our data clearly indicate a superiority effect of facial over acoustic information. Impression targets received highest trustworthiness ratings in the trustworthy face + neutral voice condition, and lowest ratings in the non-trustworthy face + neutral voice condition. From this, we can conclude that the trustworthiness information in this study was in fact conveyed via the facial cues. Our results are in line with [14, 17], indicating a superior role of facial over vocal information in nonverbal communication, and thus provide further support for [13] of strong coupling between trustworthiness information and emotional valence.

We were interested in observing the relationship between confidence ratings and perceived trustworthiness ratings in order to assess whether higher confidence leads to more extreme trustworthiness ratings. We also hoped to find out whether lower confidence ratings might coincide with relatively high trustworthiness ratings that could be interpreted as being the result of a ‘leap of faith’. Our results were not very clear in either respect. For the non-trustworthy face + neutral voice condition, we have evidence that some participants with relatively low confidence have given relatively high trustworthiness ratings and we could also see the expected effect that highly confident participants rated trustworthiness extremely low. Several explanations need to be considered. On the one hand, it is possible that what we conceptualized as a ‘leap of faith’, i.e. a form of suspending uncertainty, had already occurred when people were asked to indicate their level of confidence. This is supported by the fact that we did not press them in any way to give their ratings as quickly as possible. Participants could take their time and thus any ‘leap-of-faith’ effect would be contained in the ratings already. On the other hand, it is possible that initial trustworthiness ratings do not trigger a leap of faith yet. Only when it comes to interaction with the target and decisions to make one vulnerable or not are required, is there a need to suspend uncertainty. Thus our ratings would have been collected at a point before any observable ‘leap of faith’. Finally, with a larger and more diverse sample, especially including more participants with low confidence in their trustworthiness ratings, it might have been possible to analyze why people with the same confidence rating give different trustworthiness scores for the same target (condition). A preliminary analysis of this kind may be possible, but the sample and model are a limiting factor in this. Further research can take our approach as inspiration for observing how people develop positive (or, negative) expectations in the face of uncertainty.

Future analysis of the current dataset could focus on the outliers and present case studies of specific response patterns. It is possible that the ‘leap of faith’, in the way defined earlier in this report, is only observed in participants who tend to give extreme responses and is, therefore lost with the exclusion of statistical outliers. The present study looks only at explicit judgments of trustworthiness, and it is possible that this creates a response bias in our participants i.e., when asked repeatedly to evaluate the trustworthiness level of targets they might become

suspicious and assume that there is something non-trustworthy about the targets to begin with. Therefore, we propose another measure which could provide a more implicit measure of perceived trustworthiness that could be introduced in addition to the explicit one and could eliminate the bias reaction times [18]. The link between reaction times and confidence then could provide further empirical insights on the ‘leap of faith’ phenomenon.

6 Limitations

This study is a very first attempt at addressing the issues at hand, and as such naturally possesses a couple of limitations some related to the stimulus material, and some to the experimental design. To begin with, we recognize that the trustworthiness judgments we recorded might have been at least partially influenced by the facial geometry of the actors faces, and thus are not solely the result of the nonverbal behavior manipulation. To eliminate this possibility, one might obtain trustworthiness judgments from a different participant sample based on still photographs of the actors neutral facial expressions (e.g., screenshots from the original recordings). However, since no voice-based comparison stimulus exists, this would already imply a superior role of the face in the impression of trustworthiness. Therefore, in the current study we assumed natural variance of the actors facial features shape and configuration, much resembling real life. Given this variance, we considered it valid to assume that the trustworthiness judgments provided by our participants reflected the nonverbal behavior of the actors and any effects due to facial geometry cancelled each other out.

Further concerns might arise regarding the ecological validity of the recordings used in the present study. More precisely, it could be argued that actor portrayals represent an extreme or overexaggerated nature of the facial and vocal cues that would not have been observed in day-to-day communication. This is a common concern in research relying on actor databases, including emotion expression databases. However, there is no general consensus on what in fact constitutes “natural” behavior. Very often we intentionally try to convey a certain impression, thereby seeking to influence other people. For this we put extra emphasis on certain nonverbal behaviors. Other times, nonverbal behavior in the field is influenced by a wide range of additional factors such as mood or emotional state at that precise moment, our general frame of mind or the relationship with the communication partner. By applying standardized procedures in the laboratory, we can keep most of those factors constant (or rely on sufficient variance within the entire sample) and only manipulate the variables of interest. It is true that such conditions might result in an oversimplification of the studied phenomena, but we think that such powerful exemplars are a crucial first step in order to understand these phenomena.

One last issue related to the stimulus material touches upon the potential loss or warping of audio information during the channel alignment process, which could have resulted in the observation that the voice is less informative for judging trustworthiness. We do not exclude this possibility due to the novelty of

the alignment algorithm. However, if this argument was true, then the neutral face + neutral voice condition should have received the highest trustworthiness judgments because there the stimuli did not undergo the alignment procedure and were most strongly reflecting real-life nonverbal behavior. Clearly, as can be seen in Figure 2, this was not the case. Instead, the neutral face + neutral voice condition was perceived as equally trustworthy to two other conditions where the voice had been warped. Therefore, it is safe to assume that no systematic bias in the trustworthiness judgments in this study occurred due to the channel alignment procedure.

We would like to emphasize the fact that since the present study was in itself a pilot one, its findings are yet inconclusive and need to be interpreted with care. Future investigation in the same direction would benefit from an experimental design covering four additional cue combinations (e.g., trustworthy face + trustworthy voice, trustworthy face + non-trustworthy voice, non-trustworthy face + trustworthy voice, and non-trustworthy face + non-trustworthy voice). In addition, one might be able to more clearly isolate the relative contributions of the face and voice in a between-subject design where two groups of perceivers are presented with single-channel cues (face only and voice only) and a third group is presented with the channel combination (face + voice). In fact, we are currently conducting a follow up study with the primary goal of addressing these issues. One also needs to keep in mind, however, the benefits of the simplistic design of the initial study. For example, by always keeping one cue channel constant (neutral with respect to trustworthiness) we were able to single out the exact contribution of one or the other channel. In combinations where both channels are different than neutral, the relative contribution of each modality is less clear, because the two channels produce combined effects on the resulting trustworthiness judgment. Such combinations are mostly interesting in the context of perceivers sensitivity to cue consistency, and not so much along the lines of channels relative contribution to the trustworthiness judgment (which was the purpose of the present study).

7 Conclusion

The study contributes new evidence on visual and acoustic influences on perceived trustworthiness, with particular emphasis on first impressions. Most trust research to date has investigated perceived trustworthiness in established relationships, but first impressions may determine who establishes relationships with whom. Hence prior research may be biased toward interactions that lasted beyond the initial encounter. Our research accounts for relationships that never formed and the relative role of visual and acoustic cues in this.

The present findings may have a wide range of practical implications both in the areas of everyday communication (e.g., sending a photo might be more powerful than giving a phone call) and computer animation (e.g., we see here that the generation of trustworthy virtual agent is primarily based on facial behavior rather than acoustic characteristics, [19]). Moreover, future research

in the area with the expansion to microanalytic approaches has the potential for significant contributions to the computer-based recognition of trustworthy versus non-trustworthy facial behavior.

Finally, this study contributes exploratory insights into the possibility to observe the ‘leap of faith’ in an experimental set-up and to clarify to some extent whether the initial perception of another’s trustworthiness, on trusting behavior later on, is based on “overdrawn information” [8]. Even though the results on the leap of faith in the narrower sense are inconclusive our study extends and possibly challenges prior research by combining the question of perceived trustworthiness in the other with one’s own confidence in making such an assessment. In short, we believe the current study constitutes a stimulating transdisciplinary approach to trust research, bridging experimental psychology methodology and sociological-philosophical research questions, and promising novel research designs.

Acknowledgements. This work has been conducted with the financial support of the Research Center CoWell at Jacobs University Bremen and the School of Psychology at Cardiff University. We thank Tony Manstead for his help with the recording of the actors..

References

1. Willis, J., Todorov, A.: First Impressions: Making up Your Mind after 100ms Exposure to a Face. *Psychol. Sci* (2006) 592–598
2. Ballew, C.C., Todorov, A.: Predicting Political Elections from Rapid and Unreflective Face Judgments. *Proceedings of the National Academy of Sciences of the United States of America* **104** (2007) 17948–17953
3. Surawski, M., Osoff, E.: The Effects of Physical and Vocal Attractiveness on Impression Formation of Politicians. *Current Psychology* **25** (2006) 15–27
4. Giddens, A.: *The Consequences of Modernity*. Stanford University Press, Stanford (1990)
5. Giddens, A.: *Modernity and Self-Identity*. Polity Press, Cambridge (1991)
6. Möllering, G.: The Nature of Trust: From Georg Simmel to a Theory of Expectation, Interpretation and Suspension. *Sociology* **35** (2001) 403–420
7. Möllering, G.: *Trust: Reason, Routine, Reflexivity*. Elsevier, Oxford (2006)
8. Luhmann, N.: *Trust and Power*. Two Works by Niklas Luhmann. Wiley, Chichester (1979)
9. Lewis, J.D., Weigert, A.: Trust as a Social Reality. *Social Forces* **63** (1985) pp. 967–985
10. Krumhuber, E., Manstead, A.S.R., Cosker, D., Marshall, D., Rosin, P., Kappas, A.: Facial Dynamics as Indicators of Trustworthiness and Cooperative Behavior. *Emotion* **7** (2007) 730–735
11. Kamachi, M., Hill, H., Lander, K., Vatikiotis-Bateson, E.: Putting the Face to the Voice’: Matching Identity Across Modality. *Current Biology* **13** (2003) 1709 – 1714
12. Lachs, L., Pisoni, D.: Crossmodal Source Identification in Speech Perception. *Ecological Psychology* **16** (2004) 159–187
13. Oosterhof, N.N., Todorov, A.: The Functional Basis of Face Evaluation. *Proceedings of the National Academy of Sciences of the United States of America* **105** (2008) 11087–11092

14. Hess, U., Kappas, A., Scherer, K.R.: Multichannel Communication of Emotion: Synthetic Signal Production. In Scherer, K.R., ed.: *Facets of Emotion: Recent Research*. Erlbaum, Hillsdale, NJ (1988) 161–182
15. Aubrey, A., Kajic, V., Cingovska, I., Rosin, P., Marshall, D.: Mapping and Manipulating Facial Dynamics. In: *Proc. IEEE Conf. Automatic Face and Gesture Recognition*. (2011) 639–645
16. Jarvis, B.: *MediaLab (Version 2008.1.33)*. Empirisoft Corporation, New York (2008)
17. Kappas, A., Hess, U., Scherer, K.R.: Voice and Emotion. In Feldman, R.S., Rimé, B., eds.: *Fundamentals of Nonverbal Behavior*. Cambridge University Press, Cambridge (1991) 200–238
18. Burns, C., Conchie, S.: Measuring Implicit Trust and Automatic Attitude Activation. In Lyon, F., Möllering, G., Sanders, M., eds.: *Handbook of Research Methods on Trust*. Edward Elgar, Cheltenham (2011) 239–248
19. Van Mulken, S., André, E., Muller, J.: An Empirical Study on the Trustworthiness of Life-Like Interface Agents. In Bullinger, H., Ziegler, J., eds.: *Human Computer Interaction*. Erlbaum, Mahwah, NJ (1999) 152–156