LEE JUSSIM

Abstract = 55 words
Main text = 1,564 words
Reference = 406 words
Entire text = 2,121 words

Trustworthiness perception at zero acquaintance: consensus, accuracy, and prejudice

Jean-François Bonnefon,[a] Astrid Hopfensitz,[b] Wim De Neys [c]

[a] Toulouse School of Economics, Center for Research in Management, Toulouse, France
[b] Toulouse School of Economics, GREMAQ, Toulouse, France
[c] LaPsyDE (CNRS Unit 8240), Sorbonne - Paris Descartes University, Paris, France

Toulouse School of Economics
21 allée de Brienne, 31015 Toulouse Cedex 6
France
+33 5 61 12 85 30

jfbonnefon@gmail.com
hopfensitz@gmail.com
wim.de-neys@parisdescartes.fr

https://sites.google.com/site/jfbonnefon
http://hopfensitz.weebly.com
http://www.wdeneys.org

Abstract. Research on trustworthiness perception from faces has unfolded in a way that is strikingly reminiscent of Jussim's narrative. Jussim's analysis warns us against overemphasizing evidence about prejudice over evidence about accuracy, when both are scant; and reminds us to hold all accounts to the same standards, whether they call on societal biases or true signals.

In the conclusion of his volume, Jussim mentions recent lines of research on accuracy, and in particular the accuracy of judgments at zero acquaintance, formed from photos of strangers. This comment continues the discussion engaged in these final pages, extending Jussim's argument to recent research on trustworthiness perception. Trustworthiness perceptions are especially interesting because they play a critical role in cooperation, which is itself at the very crossroad of current research in biology, economics, and psychology.

We highlight in this commentary that research on the perception of trustworthiness has unfolded in a way that is strikingly reminiscent of Jussim's overall narrative. First, the bulk of this research has focused on consensus -- that is, whether people agree about who looks trustworthy, regardless of whether they are correct or incorrect in this assessment. Second, research on trustworthiness perceptions has emphasized their potential for social injustice over their potential accuracy. Third, it has been speculated that the potential accuracy of trustworthiness perception may be due to self-fulfilling prophecies. Before we unpack each of these three points, we want to stress that our goal is not to argue that trustworthiness perceptions are fully accurate. We actually believe that the accuracy of trustworthiness perceptions is quite limited, and that they can have untoward social consequences. We also believe, though, that there is a kernel of accuracy in trustworthiness perceptions that is of broad and substantial theoretical interest. Accordingly, we wish for the field to give it full attention. As we will illustrate, this will require researchers to avoid several pitfalls vividly described by Jussim.

A large body of research has shown that people robustly agree on who looks trustworthy and who does not (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). However, studies that established this agreement were typically silent on its accuracy. For example, one paper showed that children as young as 3 or 4 rated the trustworthiness of unknown faces in a way that was consistent with adult ratings of the same faces (Cogsdill, Todorov, Spelke, & Banaji, 2014). Because these faces were artificially constructed by a computer model, though, there was no objective measure of trustworthiness against which these judgments could be compared. This is also true of another striking study which showed that ultrafast trustworthiness ratings after 100ms exposure to a face were highly correlated with judgments delivered after unrestricted time (Willis & Todorov, 2006). Another paper showed that American and Japanese gave broadly consistent ratings of the trustworthiness of political candidates, based on their pictures (Rule et al, 2010). In this case the candidates were real persons, but the study did not attempt to correlate their actual behavior to the trust they inspired.

Obviously, the main difficulty in assessing the accuracy of trustworthiness perceptions is to obtain information about the individuals in the pictures, which can serve as a benchmark of trustworthiness. Recent research on trustworthiness perceptions started to offer at least two solutions to this challenge, one based on economic games conducted in the laboratory, and another based on naturalistic decision-making in the courtroom.

The first line of research utilizes well-known behavioral economics protocols such as public good games, prisoners' dilemmas, and trust games. The trust game in particular is well suited to capture the accuracy of trustworthiness perceptions. One variant of this game involves two players, the Investor and the Trustee. The Investor is endowed with an initial sum of $10, and can choose whether to keep that money or transfer it to the Trustee. The only information available to the Investor is a photo of the Trustee. If the Investor transfers the money, the Trustee receives the $10 plus an additional $20. The Trustee then decides whether to keep the whole $30 or to split them equally with the Investor. The players cannot communicate, will not play a second round, and are completely informed about these rules and procedures. In sum, the Investor needs to decide whether to trust the other player to split the money, and the Trustee can decide whether to honor or to abuse this trust. Accordingly, the accuracy of trustworthiness perception can be measured by comparing the decisions of the Investor to the strategies of the Trustees: an Investor would demonstrate perfect accuracy by transferring money to all Trustees whose strategy is to split the money, and not transferring any money to Trustees whose strategy is to keep the money. Several articles using this protocol showed that Investors did better than chance when deciding who to trust (e.g., Bonnefon, Hopfensitz, & De Neys, 2013; De Neys, Hopfensitz, & Bonnefon, 2013, 2015; Stirrat & Perrett, 2010). However, it should be stressed that accuracy in economics games is quite limited, since Investors rarely make more than 55% correct decisions, where random decisions would be accurate 50% of the time. More importantly, this level of accuracy is only observed with cropped pictures that eliminate all but inner facial features, and disappears with full pictures showing hairstyle and clothing (Bonnefon et al., 2013). Furthermore, while Investors can show limited levels of accuracy when making their decisions about money transfers, they show no such accuracy when explicitly asked to rate the trustworthiness of the Trustees in the pictures (Bonnefon et al., 2013). In sum, this line of research has offered some evidence for the accuracy of trustworthiness detection, but also showed that it was limited in size and subject to strong contextual restrictions.

A second line of research has emphasized legal decision-making contexts, in which pictures depict individuals who stood accused, or were convicted of a crime. This line of research is promising because it can address both the accuracy of trustworthiness perceptions (do people who engage in criminal activities look untrustworthy?) and their potential for social injustice (do people who look untrustworthy receive harsher sentences?). The available evidence is scant on both fronts, though, because very few studies actually measured perceptions of trustworthiness, as opposed to, e.g., perceptions of dangerousness. We know of two studies of accuracy, which showed that faces of criminals were judged as less trustworthy than faces of exemplary citizens, but once more to a small degree and subject to contextual restrictions (Porter, England, Juodis, ten Brinke, & Wilson, 2008; Rule, Krendl, Ivcevic, & Ambady, 2013). At the same time, we know of two articles documenting prejudiced legal decisions stemming from trustworthiness perceptions. First, Porter,

Gustaw, and ten Brinke (2010) showed that mock juries required less evidence to arrive at a guilty verdict when a defendant looked untrustworthy. Second and most recently, Wilson and Rule (in press) showed that convicted criminals who looked untrustworthy were more likely to have received a death rather than a life sentence. In sum, there are few demonstrations yet that trustworthiness perceptions in legal decision-making contexts are either accurate or noxious. In such a situation, caution is required when characterizing our state of knowledge. We should be careful, for example, not to claim yet that this line of research has robustly established the unreliable and nefarious nature of trustworthiness perceptions (Bonnefon, Hopfensitz, & De Neys, 2015; Olivola, Funk, & Todorov, 2014, Todorov, Funk, & Olivola, 2015). Jussim's analyses, though, warn us of the forceful pull to interpret scant evidence as definitive when it supports bias or prejudice, but weak when it supports accuracy.

In addition to warning us against such asymmetrical interpretations of the data, Jussim highlights problematic double standards that can crop up in discussions of accuracy, and more specifically when explaining accuracy as the result of self-fulfilling prophecies. In a recent review of inferences from faces, Todorov et al. (2015) suggested that accurate trustworthiness perceptions may not imply any biological link between morphology and behavior, but could instead reflect a self-fulfilling prophecy stemming from a societal bias. According to this account, people who have the sort of looks that societal biases associate with untrustworthiness would experience discrimination, and become less cooperative as a result, even though they were just as trustworthy to begin with. This is a promising line of thought, but one that Jussim reminds us to examine just as critically as its alternative. Consider for example these four propositions: [Strong Consensus] Because of societal biases, people show strong agreement about who looks untrustworthy; [Strong Prejudice] Because of societal biases, individuals who look untrustworthy suffer from discrimination; [Self-fulfilling prophecy] Individuals who suffer from discrimination become less trustworthy as a result; [No Accuracy] People who look untrustworthy are not actually untrustworthy. Clearly, this set of propositions is inconsistent, so one proposition must be incorrect. If we adopt the societal bias narrative, that is, [Strong Consensus] + [Strong Prejudice] + [No Accuracy], we must conclude that [Self-fulfilling prophecy] is false. We are not arguing that this is the case: rather, we call attention to Jussim's warning to scrutinize claims about self-fulfilling prophecies to the same extent that we scrutinize claims about accuracy.

In sum, recent advances in the field of trustworthiness perception at zero acquaintance show striking similarities with the research reviewed by Jussim, even though the accuracy of trustworthiness perceptions is not nearly as high as (and much more fragile than) the accuracy of the judgments considered by Jussim. Nonetheless, Jussim's warnings apply well to this developing field of research: We must remain careful not to overemphasize evidence about prejudice over evidence about accuracy, when both are scant; and we must be careful to hold all accounts to the same standards of evidence, whether they call on societal biases or true morphological signals.

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. Journal of Experimental Psychology: General, 142, 143-150.

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2015). Face-ism and kernels of truth in facial inferences. Trends in Cognitive Sciences, 19, 421-422.

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: a developmental study. Psychological Science, 25, 1132-1139.

De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2013). Low second-to-fourth digit ratio predicts indiscriminate social suspicion, not improved trustworthiness detection. Biology Letters, 9, 20130037.

De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2015). Adolescents gradually improve at detecting trustworthiness from the facial features of unknown adults. Journal of Economic Psychology, 47, 17-22.

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. Trends in Cognitive Sciences, 18, 566-570.

Porter, S., England, L., Juodis, M., ten Brinke, L., & Wilson, K. 2008. Is the face a window to the soul? Investigation of the accuracy of intuitive judgments of the trustworthiness of human faces. Canadian Journal of Behavioural Science, 40, 171–77

Porter, S., Gustaw, C., & ten Brinke, L. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. Psychology Crime & Law, 16, 477-491.

Rule, N. O., Ambady, N., Adams, R. B., Jr., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: Prediction and consensus across cultures. Journal of Personality and Social Psychology, 98, 1-15.

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. Journal of Personality and Social Psychology, 10, 409-426.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. Psychological Science, 21, 349-354.

Todorov, A., Funk, F., & Olivola, C. Y. (2015). Response to Bonnefon et al.: Limited 'kernels of truth'in facial inferences. Trends in cognitive sciences, 19, 422.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. Annu. Rev. Psychol. 66, 519-545.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after 100 ms exposure to a face. Psychological Science, 17, 592-598.

Wilson, J. P., & Rule, N. O. (in press). Facial trustworthiness predicts extreme sentencing outcomes. Psychological Science.