aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
2017, Vol. 26(4) 396–400

DOI: 10.1177/0963721416686211
www.psychologicalscience.org/CDPS
SAGE

Advances in Understanding the Detectability of Trustworthiness From the Face: Toward a Taxonomy of a Multifaceted Construct

John Paul Wilson[1] and Nicholas O. Rule[2]
[1]Department of Psychology, Montclair State University, and [2]Department of Psychology, University of Toronto

Abstract
Researchers have recently shown increasing interest in assessments of trustworthiness, devoting much attention to whether trustworthiness can be detected from a person's facial appearance. This question has been investigated along diverse behavioral dimensions, using a wide variety of targets, and with great inconsistency in results. Here, we call for greater precision in defining trustworthiness. We review various subdomains of trustworthiness perception and argue that developing a more highly specified taxonomy of trustworthiness will allow for better predictions about when trustworthiness can be judged on the basis of appearance, for more precision in estimating how accurate people are in making such judgments, and for more accurate information regarding the specific cues relevant to inferring trustworthiness in each domain.

Keywords
trustworthiness, person perception, accuracy, first impressions

Choosing to trust someone is an important decision, yet people often do so on the basis of very little information. Thus, a sizable literature has rapidly grown around the question of whether people can accurately perceive others' trustworthiness from indirect static cues, such as their facial appearance. This line of inquiry hinges on two key questions. First, do people generally agree about what a trustworthy person looks like? The answer to this question seems like a clear "yes," as many studies have reported high consensus for trustworthiness judgments (Oosterhof & Todorov, 2008; Winston, Strange, O'Doherty, & Dolan, 2002). Second, do these consensus opinions reliably predict the trustworthiness of people's actual behavior? This answer is much less clear: Some research suggests that they do not (e.g., Rule, Krendl, Ivcevic, & Ambady, 2013), whereas other work suggests that they do (e.g., Stirrat & Perrett, 2010). Knowing whether trustworthiness judgments are valid is critical, as perceptions of trustworthiness influence various important decisions—even capital-punishment sentencing decisions (Wilson & Rule, 2015, 2016). Ambiguity about what constitutes trustworthiness confuses the answer to this question. Here, we summarize the literature on trustworthiness perceptions based on facial appearance to help clarify this and to illustrate the need for greater precision in how researchers define what they consider "trustworthy" behavior.

Accuracy in Person Perception

Numerous studies have investigated the accuracy of perceptions of others' social attributes and personality characteristics. For example, people can categorize others according to sexual orientation, political affiliation, and some personality traits on the basis of photos or brief videos with better than chance accuracy (Alaei & Rule, 2016; Tackett, Herzhoff, Kushner, & Rule, 2016; Tskhay & Rule, 2013). A related focus of study, which psychologists have approached in diverse ways, concerns whether trustworthiness impressions predict behavior. Consensus trustworthiness ratings predict cooperative behavior in

Corresponding Author:
John Paul Wilson, Department of Psychology, Montclair State University, 1 Normal Ave., Montclair, NJ 07043
E-mail: wilsonjoh@montclair.edu

economic games, for instance (Stirrat & Perrett, 2010; Verplaetse, Vanneste, & Braeckman, 2007), and perceivers can differentiate violent from nonviolent criminals based on photos (Stillman, Maner, & Baumeister, 2010). Furthermore, studies on lie detection have generally reported above-chance accuracy (Bond & DePaulo, 2006). Other studies have failed to find that impressions of trustworthiness predict behavior, especially when the focus is on just static facial appearance rather than more dynamic nonverbal behaviors. Rule et al. (2013), for example, found that people who committed various crimes or were observed cheating looked no less trustworthy than non-cheaters and non-criminals (see also Zebrowitz, Voinescu, & Collins, 1996).

We argue that such disparate findings have been obtained largely because researchers have not defined trustworthiness consistently. Instead, the existing literature has characterized trustworthiness using a wide range of definitions, most of which merely imply trustworthiness. Such an expansive definition makes sense if trustworthiness judgments simply stem from overgeneralized perceptions of positive and negative emotions (Oosterhof & Todorov, 2008). But this broad conceptualization creates challenges when trying to understand whether perceived trustworthiness predicts behavior. Thus, the concept of trustworthiness will benefit from more precision.

Trustworthiness Perceptions in Specific Domains

Why is thinking of trustworthiness as a single construct problematic? One primary issue is that trustworthiness can include nearly any behavior generally considered "good," with untrustworthiness describing anything considered "bad." Within this framework, researchers have studied many behaviors: aggression and violence (Carré, McCormick, & Mondloch, 2009; Carré, Murphy, & Hariri, 2013; Rule et al., 2013; Stillman et al., 2010), criminal behavior (Porter, England, Juodis, ten Brinke, & Wilson, 2008; Rule et al., 2013), deception and lie detection (Bond & DePaulo, 2006; Hartwig & Bond, 2011), academic cheating (Geniole, Keyes, Carré, & McCormick, 2014; Rule et al., 2013), financial malfeasance (Rule et al., 2013), sexual infidelity (Rhodes, Morley, & Simmons, 2013), and selfish behavior in economic games (Stirrat & Perrett, 2010; Verplaetse et al., 2007). Although much of this work has not claimed to focus on trustworthiness, authors often measure trustworthiness alongside another chosen measure of interest (e.g., Stirrat & Perrett, 2010), trustworthiness correlates strongly with these measures (e.g., Rhodes et al., 2013), and scholars do often consider these potentially disparate topics under the same broad umbrella (e.g., Porter et al., 2008).

We therefore argue that the field should move toward a more specified taxonomy. Doing so would help to clarify which "trustworthy" behaviors people can predict from a person's appearance. Moreover, it would allow for more precise estimates of perceivers' accuracy. Finally, it would foster more specific hypotheses about the cues that predict behaviors considered trustworthy versus untrustworthy. To date, researchers have studied the perceptibility of trustworthiness across several key domains.

Cheating, honesty, and infidelity

Perhaps the most obviously valid conceptualizations of trustworthiness involve honesty and cheating. Participants in one recent study sat for a photograph and then had the opportunity to cheat on a test to win cash (Rule et al., 2013). Ratings of trustworthiness from their photos showed that cheaters looked no less trustworthy than non-cheaters; nor did these ratings relate to self-reported past cheating behavior.

Other studies examining similar constructs have focused on more stable traits related to chronic cheating or dishonesty. Zebrowitz et al. (1996), for example, found that women who behaved less honestly early in life came to look *more* honest later in life. Appearance may therefore reflect traits related to trustworthiness and honesty, but in a way that undermines accurate detection. In other research, people who looked dishonest were more likely to volunteer for research that required them to act deceptively than were people who looked honest (Bond, Berry, & Omar, 1994). Research on trait honesty has been quite limited, however, and much more research is needed.

Perhaps one of the most common everyday tests of trustworthiness involves a person's sexual infidelity. Rhodes et al. (2013) found that women could judge men's unfaithfulness from their faces but that men could not detect women's unfaithfulness. They also found that perceptions of unfaithfulness and trustworthiness were distinct, with only unfaithfulness ratings predicting infidelity. Notably, these measures relied on self-reports of infidelity, which themselves are of questionable trustworthiness (i.e., people who lie to their partners by cheating on them might also be more inclined to lie to researchers about doing so; see also Rule et al., 2013).

Economic-game behavior

Despite weak links between people's appearance and their cheating or dishonest behavior, facial information may predict how people behave in economic games. Verplaetse et al. (2007) found that people could identify the noncooperative participants in a prisoner's dilemma game with better than chance accuracy—but only from

photos taken during the decision-making moment. This result points to the potential legibility of state (but not trait) trustworthiness. In distinction, Stirrat and Perrett (2010) found that having a greater facial width-to-height ratio (fWHR) predicted a greater tendency to exploit others in a trust game, which suggests that there may be cues to trait trustworthiness. Facial structure may not relate to untrustworthy behavior in all scenarios, however. When out-group competition was salient, fWHR positively predicted cooperation with in-group members (Stirrat & Perrett, 2012). The consistency with which perceived trustworthiness predicts behavior in economic games may therefore depend on the situation.

Indeed, in some instances, only implicit trustworthiness judgments are accurate. Bonnefon, Hopfensitz, and De Neys (2013), for example, showed participants photographs of people who had previously participated in a trust game, asking the participants to play the role of trustor in a new trust game with these individuals. The participants transferred less money to partners who had failed to reciprocate in the previous game but did not rate the reciprocators and abusers differently when explicitly judging trustworthiness, showing that the domain specificity of the trustworthiness measure in question could matter in some circumstances. Amount of information also impacted these judgments: Participants accurately judged trustworthiness from photos of core facial features but not from photos that included hairstyle and clothing. The accuracy of trustworthiness judgments thus seems fragile and easily disrupted by reflective judgment processes. Shoda and McConnell (2013), for instance, observed below-chance responding among participants attempting to categorize Verplaetse et al.'s (2007) targets as defectors or cooperators. The evidence for a relationship between facial appearance and trustworthy behavior in economic games thus appears mixed.

Aggression

Physical aggression has also been linked to trustworthiness, and some evidence suggests that aggressive tendencies can be read from the face. Studies have repeatedly shown that cues such as fWHR can predict physical aggression (e.g., Carré & McCormick, 2008; Carré et al., 2009). A recent meta-analysis confirmed this link, concluding that fWHR positively predicts threat behavior in men and dominance behavior in both men and women (Geniole, Denson, Dixson, Carré, & McCormick, 2015). Most of this research has focused on objective facial measurements rather than subjective perception, but some studies have shown that perceivers can reliably detect aggressive tendencies, distinguishing aggression from other forms of untrustworthiness. In one such study, perceivers viewing mugshots of violent and nonviolent criminals accurately differentiated the two groups (Stillman et al., 2010). Thus, solid evidence suggests that the face contains cues to aggression. Moreover, people may succeed in using these accurate cues to judge aggression.

Criminality

Related to aggression (but not necessarily involving physical force), criminality has also served as a proxy for trustworthiness. In one investigation, Porter et al. (2008) found that criminals featured on *America's Most Wanted* looked less trustworthy than Nobel Peace Prize recipients. This study contrasted criminal mugshots with professional photos from the Nobel website, however, exacerbating the appearance of differences between the two groups. Indeed, Rule et al. (2013) found that trustworthiness ratings for Nobel Peace Prize winners similarly differed from those for celebrities when participants viewed mugshots of the celebrities, but not when they viewed commercial photos of the same individuals. In additional studies, they found that executives who committed financial crimes looked no less trustworthy than those who had not, and that convicted U.S. war criminals and U.S. military heroes appeared comparably trustworthy. Thus, the criminality literature seems mixed, with most recent research indicating that criminality perceptions may not be accurate and suggesting that physical aggression may be an important component of detecting criminality accurately.

Lie detection

The most extensive research related to trustworthiness perception has concerned lie detection. Researchers have investigated whether people can detect lies and deception for decades, though usually from more than just facial cues. Meta-analyses have shown that people accurately judge deception approximately 54% of the time (Bond & DePaulo, 2006). The accuracy of such judgments may remain close to chance level not because perceivers tend to use the wrong cues when attempting to detect deception, but because actual behavioral differences between liars and truth tellers are small (Hartwig & Bond, 2011). Furthermore, most studies have focused on detecting single lies, with very few examining dispositional lie detection from stable cues (but see Zebrowitz et al., 1996). Thus, the large literature on lie detection is relatively peripheral to the question of whether trustworthiness is detectable from one's appearance.

A Focus on Trustworthiness as a Trait

The lie-detection literature suggests that people can to some extent detect momentary dishonesty. Is accuracy therefore limited to state trustworthiness, or is trait trustworthiness also detectable? Some studies have shown

that people who cheat once typically cheat again (Davis & Ludvigson, 1995) and that some trustworthiness cues (e.g., fWHR) may be stable (Bond et al., 1994; Hehman, Leitner, Deegan, & Gaertner, 2013; Porter et al., 2008; Rhodes et al., 2013; Stillman et al., 2010; Stirrat & Perrett, 2010, 2012). Indeed, one recent study showed that fWHR predicted one's willingness to cheat by entering extra ballots into a lottery (Geniole, Keyes, et al., 2014). Yet Rule et al. (2013) found no relationship between perceptions of facial appearance and cheating. The seeming disconnect between these results suggests that stable facial cues may underpin some aspects of trustworthiness but that they may not be used correctly by perceivers or may not apply to untrustworthy behavior universally.

It seems reasonable that behaviors in one domain might not predict behaviors in another. For example, although high fWHR predicts less cooperation in economic games, it also predicts greater generosity toward in-group members (Stirrat & Perrett, 2012). What looks trustworthy in one situation may therefore look untrustworthy in another. Trustworthiness sometimes relies on perspective, then—contributing to its amorphous conception. For instance, people whose job involves representing the best interests of their clients, shareholders, or political constituents may engage in unscrupulous behavior to meet that goal. What appears noble to the people they represent may seem quite untrustworthy to outside observers. Despite the potential context sensitivity of the construct, researchers may benefit from considering a framework that distinguishes between behaviors that have a clear physiological basis (e.g., physical aggression, sexual infidelity) and behaviors or traits that do not (e.g., cheating, economic-game behavior, honesty). Although we recognize that this is an imperfect distinction, it aligns relatively well with the existing accuracy findings. In addition to whichever specific measures are under investigation, we recommend that researchers also systematically include measurements of trustworthiness to help move toward a meta-analytic understanding of specific versus general accuracy in trustworthiness impressions.

Conclusion

Researchers tend to use "trustworthiness" as a catchall term for behaviors characterized by warmth and positivity. In light of the reviewed research, this is usually sensible. We have variously defined trustworthiness as characterizing behaviors in terms of cooperation, honesty, lack of harmful aggression, adherence to rules, and faithfulness, among other things. However, the meaning of trustworthiness likely varies across situations. For example, a person perceived as dependable in a situation requiring intellectual diligence might seem unreliable in conflicts requiring physical formidability (e.g., Hehman, Leitner, Deegan, & Gaertner, 2015; Little, Burriss, Jones, & Roberts, 2007). On the contrary, similar to one's pet guard dog, a person who appears to be a trustworthy defender in an intergroup conflict could nevertheless turn on one and attack during a scuffle within the group (e.g., Stirrat & Perrett, 2012). Thus, trustworthiness may depend upon the level of analysis, and this contingency highlights the need for more precision in defining it.

As researchers show accelerating interest in trustworthiness perception, developing a valid taxonomy of the construct therefore seems increasingly important. This need is becoming particularly salient as evidence for the potentially dire consequences of trustworthiness perceptions continues to emerge (Wilson & Rule, 2015, 2016). With more frequent investigations of trustworthiness comes the potential for additional splintering of what it represents. This will be exacerbated as researchers continue to observe seemingly inconsistent findings that result in different operationalizations of the concept. At present, one cannot confidently answer the question of whether trustworthiness can be read from appearance. Thus, rather than adding further nuance to the answer, researchers may gain better resolution in their understanding of the phenomenon by changing how they ask the question.

Recommended Reading

Hall, J. A., Mast, M. S., & West, T. V. (Eds.). (2016). *The social psychology of perceiving others accurately*. Cambridge, UK: Cambridge University Press. An edited volume that provides a comprehensive treatment of the many facets of interpersonal accuracy.

Hartwig, M., & Bond, C. F. (2011). (See References). A meta-analysis of lie-detection literature focusing on why people fail to detect lies.

Porter, S., ten Brinke, L., & Gustaw, C. (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. An empirical demonstration of the impact of trustworthiness perceptions in the legal domain.

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). (See References). A review of the literature on the accuracy of trustworthiness judgments that provides several empirical demonstrations of lack of accuracy in such judgments.

Zebrowitz, L. A., & Montepare, J. M. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology*. New York, NY: Psychology Press. A review of the person-perception literature with an emphasis on the functional nature of perception.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

References

Alaei, R., & Rule, N. O. (2016). Accuracy of perceiving social attributes. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 125–142). Cambridge University Press.

Bond, C. F., Jr., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology*, *15*, 523–534.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*, 214–234.

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

Carré, J. M., & McCormick, C. M. (2008). In your face: Facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 2651–2656.

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, *20*, 1194–1198.

Carré, J. M., Murphy, K. R., & Hariri, A. R. (2013). What lies beneath the face of aggression? *Social Cognitive and Affective Neuroscience*, *8*, 224–229.

Davis, S. F., & Ludvigson, H. W. (1995). Additional data on academic dishonesty and a proposal for remediation. *Teaching of Psychology*, *22*, 119–121.

Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PLoS ONE*, *10*(7), e0132726.

Geniole, S. N., Keyes, A. E., Carré, J. M., & McCormick, C. M. (2014). Fearless dominance mediates the relationship between the facial width-to-height ratio and willingness to cheat. *Personality and Individual Differences*, *57*, 59–64.

Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*, 643–659.

Hehman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2013). Facial structure is indicative of explicit support for prejudicial beliefs. *Psychological Science*, *24*, 289–296.

Hehman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2015). Picking teams: When dominant facial structure is preferred. *Journal of Experimental Social Psychology*, *59*, 51–59.

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution & Human Behavior*, *28*, 18–27.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, USA*, *105*, 11087–11092.

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*, 171–177.

Rhodes, G., Morley, G., & Simmons, L. W. (2013). Women can judge sexual unfaithfulness from unfamiliar men's faces. *Biology Letters*, *9*(1), Article 20120908. Retrieved from http://rsbl.royalsocietypublishing.org/content/9/1/20120908

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*, *104*, 409–426.

Shoda, T. M., & McConnell, A. R. (2013). Interpersonal sensitivity and self-knowledge: Those chronic for trustworthiness are more accurate at detecting it in others. *Journal of Experimental Social Psychology*, *49*, 440–443.

Stillman, T. F., Maner, J. K., & Baumeister, R. F. (2010). A thin slice of violence: Distinguishing violent from nonviolent sex offenders at a glance. *Evolution & Human Behavior*, *31*, 298–303.

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

Stirrat, M., & Perrett, D. I. (2012). Face structure predicts cooperation: Men with wider faces are more generous to their in-group when out-group competition is salient. *Psychological Science*, *23*, 718–722.

Tackett, J. L., Herzhoff, K., Kushner, S. C., & Rule, N. O. (2016). Thin slices of child personality: Perceptual, situational, and behavioral contributions. *Journal of Personality and Social Psychology*, *110*, 150–166.

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review*, *17*, 72–86.

Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: The sequel: A kernel of truth in predictive cheating detection. *Evolution & Human Behavior*, *28*, 260–271.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*, 1325–1331.

Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. *Social Psychological & Personality Science*, *7*, 331–338.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*, 277–283.

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). "Wide-eyed" and "crooked-faced": Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin*, *22*, 1258–1269.