Can we detect cooperators by looking at their face?

Word count: 2,500

Jean-François Bonnefon (1)

Toulouse School of Economics, Institute for Advanced Study in Toulouse, Centre National de la Recherche Scientifique, Center for Research in Management (UMR5303), University of Toulouse Capitole, Toulouse, France

Astrid Hopfensitz

Toulouse School of Economics, Institute for Advanced Study in Toulouse University of Toulouse Capitole, Toulouse, France

Wim De Neys

LaPsyDE (CNRS Unit 8240), Sorbonne – Paris Descartes University, Paris, France

(1) Address correspondence to Jean-François Bonnefon, Toulouse School of Economics, 21 allée de Brienne, 31015 Toulouse, France, E-mail: jean-francois.bonnefon@tse-fr.eu. Support through the ANR-Labex IAST is gratefully acknowledged.

Abstract

Humans are willing to cooperate with each other for mutual benefit, and to accept the risk of exploitation. To avoid collaborating with the wrong person, people sometimes attempt to detect cooperativeness in body language, facial features, and facial expressions. But how reliable are these impressions? We review the literature on the detection of cooperativeness in economic games, from protocols that provide a lot of information on players (e.g., long personal interaction) to protocols that provide minimal information (e.g., passport-like pictures). This literature suggests that people can detect cooperativeness with some small accuracy when they interacted, or watched video clips of other players, but that they have a harder time extracting information from pictures. The conditions under which people have better than chance accuracy after seeing pictures suggests that successful cooperation detection is supported by purely intuitive processes.

Keywords: Cooperation; Detection; Face; Games; Intuition

Can we detect cooperators by looking at their face?

Humans are not very fast, not very strong, and not very intimidating. Yet they have outclassed about every other species and colonized about every corner of land. This astounding success is built upon a unique propensity to work cooperatively (Tomasello, 2014; Whiten & Erdal, 2012). All the way from coordinated hunting to building international space stations, humans are willing to cooperate with each other for mutual benefit, and to accept the risk of exploitation. Indeed, to act cooperatively is to accept vulnerability, since one can be exploited by selfish agents. Accordingly, people resort to various safeguards to protect themselves against cooperating with the wrong person (Cosmides & Tooby, 1992). These safeguards include first-hand information based on past interaction, or gossip and reputation. And in last resort, in the absence of relevant information, people can attempt to detect cooperativeness in body language, facial features, and facial expressions.

But how accurate are these impressions? Is it really possible to assess the cooperative intentions of a stranger after a brief interaction? Is it really possible to look at the picture of a stranger and make an appropriate trusting decision? The question is not whether we can perfectly assess cooperative intentions just by looking at someone. We clearly cannot, or else scammers would be out of business already. Rather, the question is one of signal to noise. Are we hopelessly inaccurate and biased cooperation detectors, or do our judgments contain a kernel of predictive power, over and beyond the accuracy that would be expected by chance alone (Bonnefon, Hopfensitz, & De Neys, 2015; Olivola, Funk, & Todorov, 2014)?

In this article, we first introduce what we believe to be the most appropriate method for studying cooperation detection: economic games in which participants can look at their partners before making a decision to cooperate. This focus has several consequences for the kind of work we will review in the rest of the article. First, we restrict our review to games in which cooperation is an applicable concept. Second, we limit our review to studies in which the detection of cooperation

is the dependent variable (as opposed to, e.g., the memory for faces of cooperators: Yamagishi, Tanida, Mashima, Shimoma, & Kanazawa, 2003). Third, we focus on whether people can detect actual cooperative tendencies, rather than on the features that make a face look more trustworthy, attractive, or dominant, independently of the diagnosticity of these features.

Although economic games were originally played under strict anonymity, variants in which players can observe each other can be used to study cooperation detection. We review the evidence for cooperation detection in such economic games, moving down from protocols that provide a lot of information about partners (e.g., long personal interaction) to protocols that provide very little information (e.g., passport-like pictures). We conclude by considering the evidence for the intuitive nature of cooperation detection from faces.

Games

A major concern for the study of cooperation detection is to use an objective, operational definition of what counts as a cooperative behavior and its detection. One solution is to use the methods of experimental economics, and to have participants play 'games' with real financial consequences, whose rules make it straightforward to characterize a decision as cooperative or not. One such game (Berg, Dickhaut, & McCabe, 1995) is depicted in Figure 1 (there are many other games which are relevant to the current survey, such as the prisoners' dilemma, the stag hunt game, or the public good game; we will review results obtained with all these games, but we do not describe them here for the sake of concision). In the trust game depicted in Figure 1, a trustee that shares the money is cooperative (i.e., displays positive reciprocity), and a trustee that keeps the money is not. If trustors were perfect cooperation detectors, they would transfer money to trustees that cooperate, and keep the money when they play with trustees that do not cooperate. Accordingly, we can measure cooperation detection by recording how accurate trustors are in their decisions, after they have had an opportunity to observe the trustee.

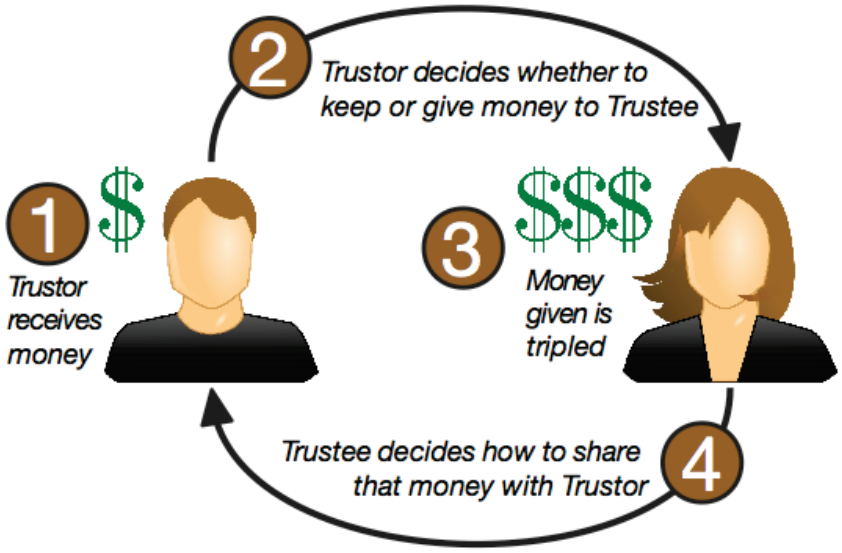

Figure 1. The trust game is used by social scientists to measure trust and reciprocity among anonymous strangers. Trustors can increase their earnings by transferring money to trustees, but trustees will face the temptation to keep the money. Accordingly, trustors should only transfer the money to trustees they expect to cooperate.

Another advantage of using economic games is that they may create especially favorable conditions for cooperation detection. First, money is at stake, which means that people have a real incentive to correctly detect cooperation—an incentive that they would miss if they were simply asked to rate how trustworthy some random strangers look. Second, the rules of the game make it acceptable and legitimate to not trust the other player—whereas in daily life, there are social costs attached to treating another person as untrustworthy. Overall, economic games provide people with a motivation to detect cooperation, and with a setting that makes it acceptable to demonstrate mistrust. These two factors improve lie detection (ten Brinke, Vohs, & Carney, 2016), and they may improve cooperation detection, too.

Interaction and movies

There is convergent evidence that if you give people ample time to interact face-to-face, say for 10–30 minutes, they can predict whether their interaction partner will cooperate, with better-than-chance accuracy. This result was originally reported by Frank, Gilovich, and Regan (1993) using the prisoner's dilemma game (Axelrod, 1984), and later reproduced in other studies (Brosig, 2002; Reed, Zeglen, & Schmidt, 2012; Sparks, Burleigh, & Barclay, 2016). Of course, a lot of information is available from 30 minutes of face-to-face interaction. You get to see body language and facial dynamics, you can see and hear cues of emotion, you have an opportunity to listen to what your partner has to say, and you can even prompt her to discuss matters that might help you assess her trustworthiness. So, it is good to know that people can extract some useful signals from this information-rich situation, but we need to ask how successful they are when they get less information.

For example, what if you can only observe someone when she is interacting with another person, without being an interaction partner yourself? Sylwester, Lyons, Buchanan, Nettle, and Roberts (2012) tried this with the prisoner's dilemma, showing participants either short or long video clips of other people interacting. 'Short' here means a few seconds, and 'long' means 2 minutes, considerably less than in the studies we considered so far. Furthermore, the people in the clips were not idly chatting, but actively trying to persuade each other that they would cooperate. Participants were slightly better than chance in the predictions they made from the short video clips, but not from the longer clips. Using a very similar procedure, Belot, Bhaskar, and Van de Ven (2012) reported better than chance detection for clips whose length varied from 6 to 25 minutes.

Now what if you can see a clip of someone who is just introducing herself, but not interacting with anyone? Vogt, Efferson, and Fehr (2013) tried this with 20s clips, that the participants watched with the sound either on or off. The sound did not make a difference—participants were slightly better than chance at detecting whether the person would cooperate in the

prisoner's dilemma (but not in the stag hunt game, Skyrms, 2004), even if they could not hear what that person said. Comparably, Fetchenhauer, Groothuis, and Pradel (2010) showed that participants were better than chance at predicting behavior in the dictator game (Engel, 2011) after watching a silent, 20s clip of the players. In sum, although no meta-analysis summarized all the evidence yet, the balance of evidence suggests that people can detect cooperation from dynamic facial displays. But what about static displays, that is, pictures?

Pictures

The evidence for cooperation detection from pictures is not entirely straightforward—in particular, and as we will see shortly, not any picture will do. To start with positive evidence, Verplaetse, Vanneste, and Braeckman (2007) reported a better-than-chance detection of cooperation (in the prisoner's dilemma) when they just showed participants a picture of another player. One concern, though, is that this result was not replicated (using the same pictures) by Sylwester et al. (2012). Another concern is that detection was accurate only when the pictures were taken at the very moment the players made the decision to cooperate or not, and not when they were taken before the game. This leaves the possibility that participants picked some subtle emotional expression linked to that decision. Ideally, we would like to know whether people can predict the cooperation of another person, from a picture of that person taken outside the context of the game.

Tognetti, Berticat, Raymond, and Faurie (2013) tried this approach, with an interesting twist. They took pictures of players of a public goods game (Fischbacher, Gächter, & Fehr, 2001) in rural Senegal, and checked whether urban French participants could detect the cooperativeness of the players based on these pictures—that is, they investigated whether cooperation detection from facial features could be achieved cross-culturally. They did find better-than-chance cooperation detection, but only for male faces, which complicates the interpretation of their results. Other studies demonstrated some degree of cooperation detection from still pictures, but the results either did not reach statistical significance (Eckel & Petrie, 2011, trust game), or could be attributed to other

pieces of information that were leaked together with the pictures (Efferson & Vogt, 2013, trust game).

Figure 2. Full pictures can be cropped at the ears, eyebrows and chin in order to only display inner features of the face. Make-up and skin tone are de-emphasized by turning the picture to grayscale.

Given these results, one could think that pictures are simply not informative enough to elicit cooperation detection. People can detect cooperation after personal interaction, or after seeing (even brief, even mute) video clips of another person—but pictures would not give them enough to work with. It may come as a surprise, then, that cooperation detection can be improved by further degrading the informational content of the pictures.

Indeed, cooperation detection springs back to better-than-chance levels (in the trust game) when pictures are cropped and grayscaled as in Figure 2 (Bonnefon, Hopfensitz, & De Neys, 2013; De Neys, Hopfensitz, & Bonnefon, 2013, 2015; Stirrat & Perrett, 2010). Why is that? One possible reason is that transformed pictures help people to focus on the signal (inner features) and ignore the noise (clothing, hairstyle, etc.). Another possible reason is that transformed pictures discourage people from thinking too much. The idea here is that just as the unconscious mind is better than the conscious mind at detecting lies (ten Brinke et al., 2016), people may be better at detecting cooperation when they rely on intuitive processing.

Consider this finding (Bonnefon et al., 2013). When people make trusting decisions based on full, non-transformed pictures, their (inaccurate) decisions are perfectly predicted by ratings of how trustworthy the other person looks (as measured by asking this exact question to an independent sample of participants). But when people make trusting decisions based on transformed pictures, their decisions get more accurate while being far less correlated by these same ratings of how the other person looks. All happens as if full pictures would prompt people to ask themselves the explicit question ‘how trustworthy does that person look?’, and lead them astray as a consequence of this conscious processing—whereas transformed pictures would discourage people to think too much, and thus encourage them to follow their intuition, with better results.

If this interpretation is correct, then we should be able to find direct evidence that (successful) cooperation detection is based on intuitive processing. We will consider such evidence in the next section.

Intuitive processing

When psychologists want to demonstrate that a behavior is driven by intuition, they can go through a standardized checklist of experimental tests, inspired by dual-process theories (De Neys & Bonnefon, 2013; Evans, 2008; Evans & Stanovich, 2013; Pennycook, Fugelsang, & Koehler, 2015; Sloman, 1996). If a decision is driven by intuition, then people with better aptitudes for deliberate thinking (e.g., general intelligence) should not be at an advantage, or might even be at a disadvantage. And if a decision is driven by intuition, then it should survive (or even improve) under experimental manipulations that prevent people from thinking deliberately—either by forcing them to process information very quickly, or by asking them to multi-task so that they cannot focus their full attention on their decision.

Are smarter people better at detecting cooperation (from transformed pictures)? The answer is no. When Bonnefon et al. (2013) gave participants a classic intelligence test (Raven's advanced progressive matrices, Bors & Stokes, 1998), they observed that participants with lower scores on the test detected cooperation just as well as participants with higher scores.

Does multi-tasking impair cooperation detection (from transformed pictures)? The answer is no again. Bonnefon et al. (2013) asked people to memorize either simple or complex patterns of dots while making their decisions. This 'dot matrix' task (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001) is useful because it allows researchers to vary the mental burden of participants from negligible (memorizing simple patterns of dots) to highly consuming (complex patterns of dots). Participants under highly consuming mental burden detected cooperation just as well as participants under negligible mental burden.

Future research will tell us whether cooperation detection survives when participants have very little time to process pictures, but the available evidence suggests that successful cooperation detection from pictures is driven by intuitive processing. Apparently, to detect cooperation in economic games, one does not need to be smart, one does not need to concentrate, and in fact one

might be better off not thinking too much. It will be important for future research to investigate the robustness of these findings as well as their boundary conditions.

Conclusion and a word of caution

Economic games are especially suitable for the study of cooperation detection: they allow for a precise operational definition of cooperation and its detection; they provide financial incentives for accurate cooperation detection; and they make it socially acceptable to act on this detection. When they play economic games, people seem to extract useful information from the observation of their partners. This is especially true if they can interact with their partners, or watch video clips of their partners.

People have a harder time, though, extracting information from static displays such as passport-like pictures. What helps in this case is to provide people with cropped pictures focusing on the inner features of the face. Under these circumstances, the available evidence suggests that purely intuitive processes support successful cooperation detection at better than chance levels.

We should conclude with a word of caution. To say that people detect cooperation better than chance when looking at other individuals, and that they do so by intuition, should not be taken as an encouragement to 'go with your gut feelings' when deciding whether someone should be trusted. The accuracy of face perception in the cooperative domain is better than chance but still very small, which means that it can be easily washed in the social biases and prejudice that pervade facial impressions (Olivola et al., 2014; Todorov, Funk, & Olivola, 2015). That people show some measure of accuracy when assessing cooperation from faces is a scientifically important phenomenon, which we must investigate and understand—but we must be just as careful not to let this finding be unduly amplified in popular media or policy recommendations.

References

Axelrod, R. (1984). *The evolution of cooperation*. NY: Basic Books.

Belot, M., Bhaskar, V., & Van de Ven, J. (2012). Can observers predict trustworthiness? *Review of Economics and Statistics*, *94*, 246–259.

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.

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, *58*, 382–398.

Brosig, J. (2002). Identifying cooperative behavior: Some experimental results in a prisoner's dilemma game. *Journal of Economic Behavior & Organization*, *47*, 275–290.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind.* NY: Oxford University Press.

De Neys, W., & Bonnefon, J. F. (2013). The whys and whens of individual differences in thinking biases. *Trends in Cognitive Sciences*, *17*, 172–178.

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*, 17–22s.

De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2016, May). *Facial trustworthiness detection: Blink or think?* Paper presented at the Human Face in Economics workshop, Toulouse, France.

Eckel, C. C., & Petrie, R. (2011). Face value. *American Economic Review*, *101*, 1497–1513.

Efferson, C., & Vogt, S. (2013). Viewing men's faces does not lead to accurate predictions of trustworthiness. *Scientific Reports*, *3*, doi 10.1038/srep01047.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics, 14,* 583–610.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning. *Annual Review of Psychology, 59,* 255–278.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*, 223–241.

Fetchenhauer, D., Groothuis, T., & Pradel, J. (2010). Not only states but traits—humans can identify permanent altruistic dispositions in 20s. *Evolution and Human Behavior*, *31*, 80–86.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economic Letters, 71,* 397–404.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: an experiment. *Ethology & Sociobiology*, *14*, 247–256.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640.

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

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72.

Reed, L., Zeglen, K., & Schmidt, K. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner's Dilemma game. *Evolution and Human Behavior*, *33*, 200–209.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure.* Cambridge: Cambridge University Press.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22.

Sparks, A., Burleigh, T., & Barclay, P. (2016). We can see inside: Accurate prediction of Prisoner's Dilemma decisions in announced games following a face-to-face interaction. *Evolution and Human Behavior*, *37*, 210–216.

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

Sylwester, K., Lyons, M., Buchanan, C., Nettle, D., & Roberts, G. (2012). The role of theory of mind in assessing cooperative intentions. *Personality and Individual Differences*, *52*, 113-117.

ten Brinke, L., Vohs, K. D., & Carney, D. R. (2016). Can ordinary people detect deception after all? *Trends in Cognitive Sciences*, *20*, 579–588.

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*, 422–423.

Tognetti, A., Berticat, C., Raymond, M., & Faurie, C. (2013). Is cooperativeness readable in static facial features? An intercultural approach. *Evolution and Human Behavior*, *34*, 427-432.

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

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 and Human Behavior*, *28*, 260–271.

Vogt, S., Efferson, C., & Fehr, E. (2013). Can we see inside? Predicting strategic behavior given limited information. *Evolution and Human Behavior*, *34*, 258–264.

Whiten, A., & Erdal, D. (2012). The human socio-cognitive niche and its evolutionary origins. *Philosophical Transactions of the Royal Society B*, *367*, 2119–2129.

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior, 24,* 290-301.

Recommended Readings

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

An experimental investigation of whether accurate trustworthiness detection is supported by intuitive processes.

DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. Psychological science, 23, 1549-1556.

An article attempting to identify nonverbal cues of trustworthiness—and to implement them in robots.

Eckel, C. C., & Petrie, R. (2011). Face value. American Economic Review, 101, 1497–1513.

A comprehensive investigation of how much people are willing to pay to see their game partner, and what happens when they do.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. Annual Review of Psychology, 66, 519-545.

A broader review of the inferences that people make from faces, with an emphasis on their unreliability.

Figure Captions

Figure 1. The trust game is used by social scientists to measure trust and reciprocity among anonymous strangers. Trustors can increase their earnings by transferring money to trustees, but trustees will face the temptation to keep the money. Accordingly, trustors should only transfer the money to trustees they expect to cooperate.

Figure 2. Full pictures can be cropped at the ears, eyebrows and chin in order to only display inner features of the face. Make-up and skin tone are de-emphasized by turning the picture to grayscale.