

University of Groningen

Not only states but traits - Humans can identify permanent altruistic dispositions in 20 s

Fetchenhauer, Detlef; Groothuis, Ton; Pradel, Julia

Published in:
 Evolution and Human Behavior

DOI:
[10.1016/j.evolhumbehav.2009.06.009](https://doi.org/10.1016/j.evolhumbehav.2009.06.009)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
 Fetchenhauer, D., Groothuis, T., & Pradel, J. (2010). Not only states but traits - Humans can identify permanent altruistic dispositions in 20 s. *Evolution and Human Behavior*, 31(2), 80-86. DOI: 10.1016/j.evolhumbehav.2009.06.009

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Original Articles

Not only states but traits — Humans can identify permanent altruistic dispositions in 20 s

Detlef Fetchenhauer^{a,b}, Ton Groothuis^c, Julia Pradel^{a,*}

^aDepartment of Economic and Social Psychology, University of Cologne, Cologne, Germany

^bInteruniversity Center for Social Science Theory and Methodology, University of Groningen, Groningen, The Netherlands

^cDepartment of Behavioral Biology, University of Groningen, Groningen, The Netherlands

Initial receipt 17 September 2008; final revision received 22 June 2009

Abstract

Humans behave altruistically in one-shot interactions under total anonymity. In search of explanations for such behavior, it has been argued that at least some individuals have a general tendency to behave altruistically independent of profitability. In fact, a stable altruistic trait would be adaptive if it were recognizable. Then, altruists could choose each other in order to retain benefits through mutual cooperation. Previous research has shown that individuals can predict the degree of altruistic behavior of strangers by reading signs of emotions evoked in significant social decisions. However, the identification of benevolent emotional states is no guarantee of the existence of permanent altruistic traits, though permanent traits are the preferable criterion for selection of good interaction partners. In this study, we tested whether individuals are able to identify altruistic traits. Judges watched 20-s silent video clips of unacquainted target persons and were asked to estimate the behavior of these target persons in a money-sharing task. As the videotapes of the target persons had been recorded in a setting unrelated to altruistic behavior, the judges could not base their estimates on situational cues related to the money-sharing task but instead had to draw on stable signals of altruism. Estimates were significantly better than chance, indicating that individuals can identify permanent altruistic traits in others. As this mechanism raises opportunities for selective interactions between altruists, our findings are discussed with respect to their relevance for explaining the evolution of altruism through assortment.

© 2010 Elsevier Inc. All rights reserved.

PsycInfo classification: 2360 Motivation and emotion; 3040 Social perception and cognition.

Keywords: Altruism; Assortment; Commitment model; Evolution; Game theory; Dictator game; Person perception; Thin slices

1. Introduction

Many studies in both industrialized (Camerer, 2003) and small-scale societies (Henrich et al., 2004) have shown that when people are given the opportunity to behave selfishly in an anonymous interaction, substantial fractions of each sample behave altruistically. This finding has been a challenging issue for evolutionary researchers, as it remains unclear how altruism, which by definition reduces personal fitness, could have survived the pressures of natural selection. While a variety of theories have been advanced to explain this finding, one axis along which such theories vary is the question of whether a stable disposition to act altruistically should be identifiable. The theories of sexual selection

(Miller, 2000, 2007), indirect reciprocity (Alexander, 1987; Nowak & Sigmund, 2005) and ongoing arms races (Frank, 1988, 2004, 2008) all presuppose that trait-level altruism is identifiable — this renders selective cooperation between high-virtue individuals possible and explains why altruism eventually might turn out to be adaptive. Extending prior research on this issue, this study investigates whether trait-level altruism is indeed identifiable.

Why should altruism be identifiable? Referring to Miller (2000, 2007), altruism as well as other moral virtues evolved as fitness indicators for mate choice due to sexual selection. Miller assumes that altruism entails two types of signaling functions: on the one hand, altruism advertises an individual's fitness in terms of parenting- and relationship-coordination abilities; on the other hand, altruism promotes good genes as it depends on empathetic social intelligence, which is absent if one has a high mutation load that impairs

* Corresponding author.

E-mail address: julia.pradel@uni-koeln.de (J. Pradel).

the development of a complex Theory of Mind. According to costly signaling theory (Gintis, Smith, & Bowles, 2001; see McAndrew, 2002 for review), altruism is thus a profitable, hard-to-fake signal that may help to acquire mates and can remain reliable over time (Zahavi & Zahavi, 1997).

Helping is not only sexually attractive but also vital for a well-functioning society. In his theory of indirect reciprocity, Alexander (1987) hypothesizes that altruism, which advertises an inherent tendency to cooperate, attracts not only mates but compensations from third parties. To single out those individuals who deserve indirect reciprocity, third parties acquire information about a person's altruism in two ways: by directly observing a person or by obtaining information about an individual's past behavior from someone else. Both of these methods imply that third parties are able to identify trait-level altruism in others.

In his commitment model, the economist Robert H. Frank suggests a manner in which the direct identification of altruists might work (Frank, 1988, 2004, 2008). He postulates that altruists are proximately driven by strong moral emotions about fairness and shame, even if they are unobserved. As signs of these emotions are visible, altruists may identify each other and cluster in teams. By means of cooperation, they may yield extraordinary gains so that finally they outcompete egoists. With reservations, Frank has been the first to explicitly predict that individuals behave altruistically even when they are on their own. Indeed, if others were really able to identify such unconditional altruists, this mechanism could explain the continuing variation in altruistic tendencies among humans. On the one hand, the higher the percentage of altruists in a given population, the less monitoring of the altruism of others takes place, and the more adaptive it is to act nonaltruistically. On the other hand, the lower the percentage of altruists in a given population, the more monitoring takes place, and the more adaptive it is to act altruistically. This leads to frequency dependent selection and an equilibrium of altruists and nonaltruists existing side by side. There is one obvious criticism of this argument: if signaling altruism is profitable, natural selection should create a deceptive copy of the altruistic signal (Fehr & Fischbacher, 2005). Though this argument is pressing, it can be reversed, as the existence of a deceptive copy should lead to the modification of the original signal in order to prevent plagiarism (Frank, 2005). An arms race between the true signal and its fraudulent copy should arise. For the present moment of our evolutionary history, the relative strengths of the true signals of altruism are unknown, and it remains an open question whether and how far humans may signal their altruistic dispositions unmistakably so that others can reliably identify them.

Previous research revealed the first evidence for the identifiability of altruism. Pradel, Euler, and Fetchenhauer (2009) showed that school students are able to estimate the dictator game decisions of their classmates better than chance. Frank, Gilovich, and Regan (1993) and Brosig (2002) showed that people can predict how others behave in

an experimental game after being acquainted for only half an hour. Other research hints at the spontaneous identifiability of altruism. On the basis of pictures, Yamagishi, Tanida, Mashima, Shimoma, and Kanazawa (2003) asked observers to memorize the faces of unknown target persons who had played prisoner's dilemma games earlier. Observers were better able to remember the faces of cheaters than the faces of cooperators. In a similar way, Verplaetse, Vanneste, and Braeckman (2007) asked observers to identify targets who had played cooperatively in such a game. Observers succeeded, but only if they responded to event-related pictures that were taken during the decision-making moment rather than when they responded to neutral pictures. The authors concluded that faces are not untrustworthy in themselves, yet, in challenging social settings, they show observable cues of noncooperativeness.

Summarizing these results, humans seem to have a cheater detection module. However, in all of these studies, participants observed targets either in or directly before the moment when those persons made the social decision in question. Observers thus could base their estimates on cues to the emotional states of the targets. It thus remains unclear whether individuals can identify stable altruistic traits. However, the theories of sexual selection (Miller, 2000, 2007), indirect reciprocity (Alexander, 1987; Nowak & Sigmund, 2005) and ongoing arms races (Frank, 1988) hypothesize exactly this — that stable altruistic traits are identifiable. Therefore, in the present study, we specifically address this premise.

We created event-unrelated stimulus material on the basis of the so-called thin slices paradigm (Ambady, Bernieri, & Richeson, 2000; Ambady & Rosenthal, 1993). This paradigm has been used to show that people are able to identify permanent characteristics of others, e.g., intelligence (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004), sociosexuality (Gangestad, Simpson, DiGeronimo, & Biek, 1992) or personality disorders (Oltmanns, Friedman, Fiedler, & Turkheimer, 2004), by watching short videotapes of target persons. In a first step, we created silent video clips of target persons lasting 20 s; afterwards, to measure their level of altruism, we asked them to play a dictator game. In this paradigm, two persons interact with each other only once and under conditions of total anonymity. The "dictator" is given a certain amount of money and has to divide this money between himself or herself and an unknown recipient; the recipient has no means of influencing or vetoing the decision of the unknown dictator. In a second step, we asked judges to watch the videos and to rate how the targets had behaved in the dictator game. Since we investigated ratings that were based on stimulus material that was completely unrelated to altruistic behavior and that was recorded before the dictator game took place, we could determine whether individuals were able to discern permanent cues to altruism.

Brown, Palameta, and Moore (2003) previously used the thin slices paradigm to test whether self-reported altruism is detectable by first impression alone. Judges' ratings

correlated significantly with the self-reported altruism of target persons. But how expressive is this finding? Although personality psychologists typically measure prosocial personality dispositions by means of self-report scales (e.g., the dimension of agreeableness in the NEO-FFI by Costa & McCrae, 1998), from an evolutionary perspective this approach bears weaknesses. According to the signaling theory, egoists should be motivated to disguise their true character. Therefore, subjects in self-reports must be expected to deceive by stating an exaggerated level of altruism. In fact, Brown et al. (2003) controlled for deceptive responding by showing that the altruism scale that they used was not significantly correlated with a measure of social desirability. However, deception may even generate self-deception, as positive illusions facilitate hiding deception from others; such deception can be achieved with modest cognitive cost (Trivers, 1991). Thus, biased self-reports may result unintentionally.

To avoid effects of deception and self-deception, this study followed previous cheater detection studies (Brosig, 2002; Frank et al., 1993; Pradel et al., 2009; Verplaetse et al., 2007; Yamagishi et al., 2003) by assessing the altruistic tendencies of target persons via a behavioral measure — the dictator game. Compared to other experimental games (e.g., the Prisoner's dilemma game), the dictator game, in our view, represents the preferable instrument, not only because it continuously assesses actual prosocial behavior but also because the specific characteristics of the paradigm, namely, noniteration and anonymity, provide no strategic reasons to behave altruistically at all.

Admittedly, behavioral measures, just like self-reports, bear weaknesses. Critics argue that the dictator game entails substantial demand characteristics (e.g., Bardsley, 2008) and that dictator game giving does not always reflect a concern for others' welfare per se but rather the attempt not to violate others' expectations (Dana, Cain, & Dawes, 2006). However, these arguments do not alter the fact that individuals vary in their degree of dictator contributions. So, even if dictator game giving is not purely altruistically motivated but is to some extent distorted by an individual's propensity to conform to prosocial norms, it nevertheless reflects an individual's tendency to cooperate under situations of scrutiny. Going a step further, one might ask whether an experimental game as a single measure is favorable to assess dispositional altruism, which is defined as the general tendency to behave altruistically in a variety of situations and on different occasions. However, we believe that dictator game behavior is a valid indicator of dispositional altruism for three reasons: (1) Experimental game behavior is stable over time. With a delay of 3 weeks, Fetchenhauer (unpublished data) played repeated dictator games with individuals and found correlations of $r=.50$. (2) Experimental game behavior is affected by genes. Wallace et al. (2007) asked twins to play an ultimatum game for real monetary stakes and showed that more than 40% of the variation in subjects' rejection

behavior was explained by additive genetic effects. (3) Dictator game behavior correlates significantly with justice sensitivity (Fetchenhauer and Huang, 2004); this personality construct, in turn, has been proven to predict prosocial tendencies in various domains (e.g., Gollwitzer, Schmitt, Schalke, Maes, & Baer, 2005).

To summarize, we assume that altruistic dictator game behavior may be attributed to a general, genetically determined preference of the actor to display prosocial acts. Asking judges to predict the dictator game behavior of target persons on the basis of neutral video clips that showed targets in a situation unrelated to altruistic behavior thus gave us the opportunity to test the hypothesis that trait-level altruism is identifiable by first impression.

2. Methods

In Step 1, we invited students ($N=56$; 26 females, 30 males) of business administration to the laboratory, where they were videotaped sitting behind a desk and in front of a white wall. The target persons were asked to talk into the camera and briefly introduce themselves. For each target person, the zoom was calibrated in such a way that the face and the upper body could be seen in the video.

After being videotaped, the target persons were escorted to a different room where they were left alone to fill in a questionnaire uninterrupted. Target persons initially completed three personality scales, which served as distraction tasks. Afterwards, via the instructions, target persons were familiarized with the logic of the dictator game without naming the game as such. Target persons were asked to imagine a situation with two individuals, a dictator and a recipient, in which the dictator receives a certain amount of money. Taking the perspective of the dictator, participants were required to consider the distribution of the money between themselves and an unknown recipient. To avoid influencing the participants in any possible way, neutral vocabulary was used throughout the instructions. For example, the dictator game was referred to as a "distribution task," the dictator was referred to as "Person A," and the recipient was named "Person B."

Target persons were given €60 (about US\$94) and had four alternatives for dividing the money: they could give either €30, €20, €10 or nothing to the recipient. By a random mechanism, one out of six of them was making the decision for real. They were also informed that their interaction partner was another randomly chosen individual who participated in the study, but that this recipient would never be informed about the identity of the dictator who was accountable for his or her outcome. The students were asked to write a code word on their questionnaire and to put the questionnaire into a sealed envelope. A week after the experiment, participants could get their money from the secretary of the department. Participants stated their code-word and received an envelope with the money inside.

On average, the targets gave €20.17 (S.D., ±11.67) to the recipient. This corresponds to a proportion of 33.6% of the money at stake, which is in line with mean contributions found in former dictator game studies (Camerer, 2003). The frequency of fair dictators was slightly higher than in most other studies, as 27 targets (48.2%) split the money equally (i.e., gave €30 to the recipient). Additionally, 14 (25.0%) targets gave €20 to the recipient and four (7.1%) of them gave €10 to the recipient. A substantial minority of 11 participants (19.6%) decided to transfer nothing but to keep everything for themselves.

In Step 2, the logic of the dictator game was explained to a group of biology students ($N=34$; 18 females, 16 males) of the University of Groningen. These judges were then shown the videos of the 56 different target persons. Each video lasted 20 s, and the sound was switched off during the whole experiment. After each video, the judges had to estimate which of the four alternatives the target had chosen in the dictator game. Under no circumstances did the judges receive feedback on the correctness of their estimates. Moreover, on inquiry, it was assured that none of the judges knew any target persons.

3. Results

Our hypothesis was that judges are able to rate the altruism of unknown target persons better than chance. To test this, we initially investigated how closely the average estimates of judges matched the actual dictator contributions of the targets. To accomplish this, we first combined the estimates of the 34 judges into one single scale (Cronbach's $\alpha=.84$; interrater correlation=.14) and then correlated this scale with the actual dictator game contributions. If the

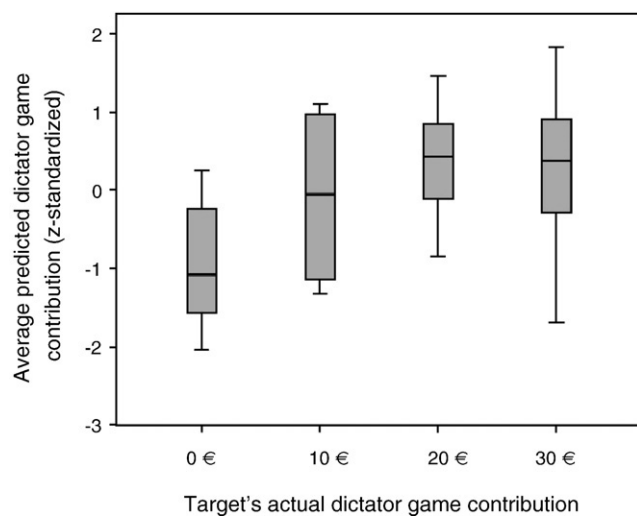


Fig. 1. Relationship between the actual dictator game contributions of target persons and the mean estimates of judges concerning the targets' contributions. Estimates for those who contributed nothing differed significantly from estimates for those who contributed €10, €20 or €30.

Table 1

Association between the actual dictator game contributions of targets and the corresponding frequency of estimates

Targets' actual dictator game contributions	Frequency of estimated dictator game contributions			
	0	10	20	30
0	31%	26%	28%	15%
10	21%	29%	21%	29%
20	16%	23%	35%	26%
30	16%	24%	28%	32%

ratings of judges had been random, a null correlation would have resulted. However, actual and average estimated dictator game contributions correlated significantly ($r=.41$, $p<.01$), indicating that the average estimates of judges were remarkably accurate. Giving z -standardized values, Fig. 1 highlights this result. As can be seen, the average estimated contribution pertaining to a person who kept all the money was more than 1 S.D. lower than the average predicted contribution of a person who divided the money equally (i.e., who transferred €30). Post hoc tests revealed that average estimates for those who acted completely self-interestedly and transferred nothing differed significantly from the average estimates for those in the other three categories ($p<.05$). However, differences between the average estimates for the three other categories (i.e., subjects giving €10, €20 or €30 to the recipient) were not significant. Fig. 1 thus illustrates that the 34 judges together performed particularly well in identifying those target persons who had kept all the money for themselves.

Admittedly, averaging estimates levels out outliers. For more strictness, we thus calculated individual accuracy rates for all 34 judges by correlating their individual estimates with the actual contributions of targets. The lowest of the resulting 34 correlation coefficients amounted to $r=-.09$ (ns), while the highest amounted to $r=.33$ ($p<.01$). The individual accuracy rate was negative for only three out of the 34 judges. Averaging individual accuracy rates led to a mean correlation coefficient of .16 (S.D., ±.11), which was significantly different from zero (one-sample t test, $t=8.936$; $p<.001$). There was no difference in the validity of estimates made by male or female judges. The average value for male judges was .16 (S.D., ±.11); the average value for female judges was .17 (S.D., ±.11). A t test revealed that this difference was not significant ($t=.274$, ns).

To highlight the accuracy of individual estimates, Table 1 shows the relationship between stimulus persons' actual dictator contributions and the frequency of different estimates about these decisions. With regard to those stimulus persons who actually contributed nothing, 31% of all judges thought that they had given nothing to Person B, while only 16% of all judges thought that they had split the money evenly. With regard to those stimulus persons who actually split the money by contributing €30, 32% of all judges estimated that they had split the money evenly; only

15% thought they had given nothing to Person B. In fact, for each value of stimulus persons' actual behavior, judges exactly estimated this value most often.

But which cues did the judges use for their estimates? One valid cue could have been the sex of the target, as females on average gave away more money than males [mean=€24.61 vs. mean=€16.33; $t(56)=2.81$, $p=.007$]. In fact, judges took this sex difference into account by assuming that female targets had given away more money (mean=€18.67) than males (mean=€14.10) [$t(56)=3.77$, $p<.001$]. Nevertheless, the accuracy of the judges' estimates cannot be explained by the identification of the sex difference alone. A regression analysis using judges' average estimates as the dependent variable and both the targets' sex and their actual behavior as independent variables showed that both independent variables were significant predictors: $F(2)=9.798$ ($p<.01$). The impact of sex was $\beta=.34$ ($p<.01$), whereas the impact of the targets' actual behavior amounted to $\beta=.29$ ($p<.05$).

4. Discussion

The theories of sexual selection (Miller, 2000, 2007), indirect reciprocity (Alexander, 1987) and ongoing arms races (Frank, 1988, 2004, 2008) hypothesize that some individuals pursue ultimate genetic self-interest through psychological adaptations that embody a genuine concern for others. If such individuals carry observable cues of their altruistic traits, these cues should be used by others in order to accurately assess strangers' dispositions.

Our study aimed at testing whether individuals are able to estimate the level of altruism of unacquainted target persons on the basis of watching 20-s silent video clips. As target persons had been videotaped in a setting completely unrelated to altruistic behavior, it could be determined whether individuals were able to make permanent cues to situation-independent altruistic traits. Results showed that individuals were indeed able to estimate the altruistic behavior of target persons better than chance.

The proper identification of altruists can be advantageous in a variety of situations. In line with Miller (2000, 2007), the reliable detection of moral virtues should lead to improved assortative mating so that high virtue mates choose one another and low virtue mates have no choice but to stay among themselves. Following Frank (1988, 2004, 2008), the ability to identify altruistic tendencies might contribute to solving principal–agent problems (Grossman & Hart, 1983) in a multiplicity of relationships. Principal–agent problems arise when individuals have to choose an interaction partner but lack information about the trustworthiness and potential future behavior of this person. This problem exists in particular when the potential interaction partner is completely unknown to the individual and can be supposed to have little commitment to behave trustworthily (Schelling, 1960).

Only recently, Verplaetse et al. (2007) showed that individuals indeed have some capability to differentiate between cooperative and noncooperative individuals. However, this effect could only be evidenced when individuals responded to pictures that had been taken during the moment of prosocial decision making; it did not occur when subjects responded to neutral pictures unrelated to the decision task. Against the background of research on the thin slices paradigm, it seems surprising that Verplaetse et al. (2007) could find evidence for any cheater detection abilities at all, since, as Ambady et al. (2000) stated, short excerpts of dynamic behavior — often no longer than 30 s — enable accurate judgments of other persons, but a still photograph is not sufficient as it bears no dynamic information about the target person. Accordingly, discrepancies between the findings in hand and those of Verplaetse et al. (2007) might result from differences in the stimulus material. While Verplaetse et al. used motionless pictures, accurate assessments of permanent cues to altruism in this study were evidenced on the basis of short excerpts of dynamic behavior. In our view, the present approach is valuable in terms of external validity because our ancestors obviously did not use photographs either when judging a stranger on the basis of first impressions. However, it must be conceded that videos provide 2D information much like a photograph. Because physical features are still present, one cannot be sure that the raters' assessments are based solely on nonverbal behavior. In fact, this is a weakness of all work that has been done testing the altruist-detection hypothesis. A means of addressing this problem in future research could be to use motion-capture stimuli of movements, separating the signals from the signalers.

The altruistic tendencies of our target persons were assessed via the dictator game. Due to its potential demand, it could be that the observed dictator game giving was not purely altruistically motivated but to some extent distorted by the propensity of individuals to conform to prosocial norms. However, it can be assumed that our judges were themselves aware of the norms that would influence participants in the dictator game. Therefore, even if judges did not evaluate noncontingent altruism per se, they were able to estimate the behavior of targets under situations of scrutiny. This can be considered adaptive because altruistic persons should always — irrespective of their intrinsic or extrinsic motivation — be favored over egoists as cooperation partners.

Although the judges' estimates were far from perfect, our data indicate that university students in Western Europe are able to predict accurately the extent of altruism of others based on a limited set of dynamic cues in a short time span. Judges performed especially well in identifying those target persons who kept all the money for themselves. This result strikingly replicates findings of Pradel et al. (2009). In that study, school students were required to assess the dictator game behavior of classmates and were also most accurate in estimating decisions of selfishly playing peers. This seems reasonable from an evolutionary point of view. It is not that

important to distinguish the nice ones from the very nice ones, but it is vital to identify the real egoists, as not doing so can literally be lethal.

Our results suggest that European students honestly signal altruistic dispositions through cues other than their altruistic acts and that variation in these dispositions exists. This adds a new dimension to the study of the evolution of human altruism; as altruism is identifiable, it might simply be advantageous to behave altruistically, due to the opportunity of altruists to carefully choose like-minded individuals for mutual cooperation. In fact, when allowed to choose their partners freely, individuals try to interact with those whom they expect to cooperate more (Johnson, Price, & Takezawa, 2008). In this vein, altruistic subjects in a public good game by Page, Putterman, and Unel (2005) positively assorted when information about the contribution histories of players was given and proceeded to interact more productively than less cooperative coplayers did. Similarly, Sheldon, Sheldon, and Osbaldiston (2000) asked university freshmen to recruit three peers to participate in an N-person prisoner's dilemma game and found that prosocial individuals tended to stick together, achieving higher group-level returns than antisocial participants. Transcending settings in the laboratory, the assortment of individuals with comparable altruistic tendencies even seems to be of importance in real life. Pradel et al. (2009) observed that students who play altruistically in a dictator game choose friends who behave similarly to themselves.

As this study shows that altruistic traits are already sensed after a very short time span, it may be possible that altruists not only become choosy after a phase of shared experiences, but also that interactions are selective from the very first moment of acquaintance.

Our findings moreover open a new avenue of research on the actual cues involved in nonverbal signaling of altruistic dispositions and raise the question of why these signals are reliable and thus evolutionarily stable. One cue that judges in this study used was sex. This cue was valid as, in line with earlier studies (Eckel & Grossman, 1998), females indeed acted more altruistically than males did. However, as accuracy in judgments could not be explained by sex differences alone, other cues are likely to be relevant. One possibility is that the cues involved are involuntary facial expressions that are difficult to voluntarily control for strategic use. Even if dishonest use of these cues were possible, it might be maladaptive to use them in this way because humans' willingness to engage in altruistic punishment might be especially strong towards those who signal a high level of altruism — and then do not comply with these signals.

Acknowledgments

We would like to thank Redouan Bshary, Serge Daan, Dave Dunning, Tim Fawcett, Robb Willer and Mareike Hoffmann for their thoughtful remarks.

References

- Alexander, R. D. (1987). *The biology of moral systems*. Aldine, NY: Hawthorne.
- Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behaviour: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, 32, 201–271.
- Ambady, N., & Rosenthal, R. (1993). Half a minute: predicting teacher evaluations from thin slices of nonverbal behaviour and physical attractiveness. *Journal of Personality and Social Psychology*, 64, 431–441.
- Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, 11, 122–133.
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behaviour as cues of personality and intelligence. *Journal of Personality and Social Psychology*, 86, 599–614.
- Brosig, J. (2002). Identifying cooperative behaviour: some experimental results in a prisoner's dilemma game. *Journal of Economic Behaviour & Organization*, 47, 275–290.
- Brown, W. M., Palameta, B., & Moore, C. (2003). Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evolutionary Psychology*, 1, 42–69.
- Camerer, C. (2003). *Behavioral game theory: experiments on strategic interaction*. Princeton, NJ: University Press.
- Costa Jr, P. T., & McCrae, R. R. (1998). *The NEO-PI/NEO-FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.
- Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100, 193–201.
- Eckel, C., & Grossman, P. (1998). Are women less selfish than men? Evidence from dictator games. *Economic Journal*, 108, 726–735.
- Fehr, E., & Fischbacher, U. (2005). Altruists with green beards. *Analyse & Kritik*, 27, 73–84.
- Fetchenhauer, D., & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences*, 36, 1015–1029.
- Frank, R. H. (1988). *Passions within reasons: the strategic role of the emotions*. New York: W.W. Norton & Company.
- Frank, R. H. (2004). *What price the moral high ground?* Princeton, NJ: Princeton University Press; 2004.
- Frank, R. H. (2005). Altruists with green beards: still kicking? *Analyse & Kritik*, 27, 85–96.
- Frank, R. H. (2008). On the evolution of moral sentiments. In: Crawford C, Krebs D (Eds.), *Foundations of evolutionary psychology* (pp. 371–379). New York: Lawrence Erlbaum Associates.
- Frank, R. H., Gilovich, T., & Regan, D. (1993). The evolution of one-shot cooperation: an experiment. *Ethology and Sociobiology*, 14, 247–256.
- Gangestad, S. W., Simpson, J. A., DiGeronimo, K., & Biek, M. (1992). Differential accuracy in person perception across traits: examination of a functional hypothesis. *Journal of Personality and Social Psychology*, 62, 688–698.
- Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of Theoretical Biology*, 213, 103–119.
- Gollwitzer, M., Schmitt, M., Schalke, R., Maes, J., & Baer, A. (2005). Asymmetrical effects of justice sensitivity perspectives on prosocial and antisocial behavior. *Social Justice Research*, 18, 183–201.
- Grossman, S. J., & Hart, O. (1983). An analysis of the principal agent problem. *Econometrica*, 51, 7–46.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (2004). *Foundations of human sociality*. Oxford: University Press.
- Johnson, D. D. P., Price, E., & Takezawa, M. (2008). Renaissance of the individual. Reciprocity, positive assortment, and the puzzle of human cooperation. In: Crawford C, Krebs D (Eds.), *Foundations of evolutionary psychology* (pp. 331–352). New York: Lawrence Erlbaum Associates.

- McAndrew, F. T. (2002). New evolutionary perspectives on altruism: multilevel-selection and costly-signaling theories. *Current Directions in Psychological Science*, 11, 79–82.
- Miller, G. F. (2000). *The mating mind*. London: Vintage.
- Miller, G. F. (2007). Sexual selection for moral virtues. *Quarterly Review of Biology*, 82, 97–125.
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437, 1291–1298.
- Oltmanns, T. F., Friedman, J., Fiedler, E. R., & Turkheimer, E. (2004). Perceptions of people with personality disorders based on thin slices of behaviour. *Journal of Research in Personality*, 38, 216–229.
- Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public good experiments: reciprocity, mimicry, and efficiency. *Economic Journal*, 115, 1032–1053.
- Pradel, J., Euler, H., & Fetchenhauer, D. (2009). Spotting altruistic dictator game players and mingling with them: the elective assortment of classmates. *Evolution and Human Behavior*, 30, 103–113.
- Schelling, T. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Sheldon, K. M., Sheldon, M. S., & Osbaldiston, R. (2000). Prosocial values and group-assortation within an N-person prisoner's dilemma. *Human Nature*, 11, 387–404.
- Trivers, R. L. (1991). Deceit and self-deception: the relationship between communication and consciousness. In: Robinson M, Tiger L (Eds.), *Man and beast revisited* (pp. 175–191). Washington, DC: Smithsonian.
- 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.
- Wallace, B., Cesarini, D., Lichtenstein, P., & Johannesson, M. (2007). Heritability of ultimatum game responder behavior. *Proceedings of the National Academy of Science USA*, 104, 15631–15634.
- 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.
- Zahavi, A., & Zahavi, A. (1997). *The handicap principle: a missing piece of Darwin's puzzle*. New York: Oxford University Press.