ELSEVIER

Evolution and Human Behavior 21 (2000) 25–37

Evolution and Human Behavior

Is prospective altruist-detection an evolved solution to the adaptive problem of subtle cheating in cooperative ventures? Supportive evidence using the Wason selection task

William Michael Brown, M.Sc.*, Chris Moore, Ph.D.

Department of Psychology, Life Sciences Centre, Dalhousie University, Halifax, Nova Scotia, Canada

Received September 12, 1998; revised manuscript July 8, 1999

Abstract

Reciprocal altruism in humans may be made possible in part by the existence of information processing mechanisms for the detection of overt cheating. However, cheating may not always be readily detectable due to the division of labor. Subtle cheating poses a serious problem for the evolution of altruism. This article argues that subtle cheating may have exerted selective pressures on early hominids to be sensitive to information regarding the genuineness of an altruistic act. In two experiments, subjects were required to complete Wason selection tasks designed to allow for the detection of altruism. Performance on the altruist-detection tasks was compared to performance on control Wason selection tasks (Experiment 1) and to performance on control and cheater detection tasks (Experiment 2). Participants were significantly better at solving cheater-detection and altruist-detection versions compared to control versions of the problems, and there was no significant difference between altruist-detection and cheater-detection. Results are discussed in relation to recent conceptual models for the evolution of altruism. Specifically, it is argued that non-kin altruism may be an evolutionarily stable strategy if altruists can detect one another and form mutually beneficial social support networks. © 2000 Elsevier Science Inc.

Keywords: Prospective altruist-detection; Evolutionary biology; Subtle cheating; Reciprocal altruism; Game theory.

Evolutionary biology defines altruism as a behavior that apparently reduces the fitness of the donor and increases the fitness of the recipient (Hamilton, 1971; Trivers, 1971). The prevalence of altruism among non-kin has presented a paradox for evolutionary theory. In order for altruistic behaviour to evolve, there must be some benefit to the genes that predispose the donor to perform self-sacrificial acts. Trivers (1971) suggested that the performance of an altruistic act may increase the likelihood that the recipient reciprocates the altruism in the fu-

* Corresponding author. E-mail: wmbrown@is2.dal.ca.

PII: S1090-5138(99)00018-5

ture. Reciprocal altruism is likely to evolve when withholding reciprocity is detected and punished, thus shielding the altruist from free-rider exploitation.

Mathematical and computer simulation work using the Prisoner's Dilemma game suggest that reciprocity can be evolutionarily stable in an ecology with free riders. Specifically, in the iterated Prisoner's Dilemma tournament reported by Axelrod and Hamilton (1981), the best strategy was Tit-for-Tat.[1] Tit-for-Tat calls for cooperation on the first move and the copying of each subsequent move made by its opponent.

Strategies that retaliate against defection (e.g., Tit-for-Tat) can only evolve in an ecology where cheating is detectable. Therefore, a clear prediction from reciprocal altruism theory is that humans should possess specialized information processing capabilities to detect cheaters (Cosmides, 1989; Cosmides and Tooby, 1992). In support of this prediction, Cosmides (1989) and Cosmides and Tooby (1992) found evidence for the existence of a cheater-detection *Darwinian Algorithm* using the Wason selection task. The Wason selection task asks a subject to see whether a conditional statement of the form "If P, then Q" has been violated by any of four instances represented by four cards (Wason, 1983). Cosmides and Tooby (1992) used the Wason task to demonstrate the existence of an evolved cognitive adaptation to detect cheaters in social contracts by showing what cognitive mechanisms humans are lacking (e.g., humans cannot solve logic word problems unless they are worded in such a way where a subject is asked to detect a cheater in a social exchange situation). Detecting a cheater requires a person to look for the individual who takes the benefit without paying the cost (Cosmides and Tooby, 1992).

Although much of the evidence from human studies conforms to reciprocal altruism theory, some forms of cooperation are less obviously explained (Palameta and Brown, 1999). For example, 50% of individuals cooperated in a Prisoner's Dilemma-like game even when told they would be playing a randomly selected counterpart only once (McCabe et al., 1996). The fact that some individuals cooperate even when cheating cannot be punished suggests that human decision-making is not based on simple rules of self-interest maximization (Palameta and Brown, 1999). Unlike a Tit-for-Tat computer strategy, humans can be provoked by prosocial emotions (e.g., concern for others, guilt, and sympathy) to behave cooperatively (Frank, 1988).

It is easy to see how prosocial emotions would be at a disadvantage due to subtle cheating in division of labor situations and over evolutionary time disappear from the population. What is needed is a mechanism that shields genuine altruists from free-rider exploitation. In order for altruists to pair up and receive the maximum mutual payoff, they need to *reliably* decode their partner's intentions. Presumably, individuals whose altruism is motivated by genuine concern for others would be more likely to behave altruistically in the future (Frank, 1988; Frank et al., 1993; Hirshleifer, 1987; Trivers, 1971). *Prospective altruist-detection* could be a possible mechanism to solve the adaptive problem of subtle cheating. *Prospective altruist-detection* is the reliable detection of genuinely altruistic intentions before you enter a cooperative venture. A difference between a Tit-for-Tat strategy and prospectively detecting

[1]However, Nowak and Sigmund (1993) demonstrated that Tit-for-Tat is not always an evolutionarily stable strategy. Furthermore, Dugatkin (1997) pointed out that a mixed-strategy is always capable of invading any pure strategy—including Tit-for-Tat.

altruists is that Tit-for-Tat interacts randomly and as a result detects cheaters after it gets exploited. Unlike Tit-for-Tat, *prospective altruist-detection* allows nonrandom partner preference and therefore excludes interaction with cheaters before exploitation. Partner preference is ecologically realistic and allows for the evolution of cooperation in computer simulations (Cooper and Wallace, 1998).

To test the hypothesis that humans have information processing capabilities designed by natural selection to detect whether a helpful act was motivated by genuine concern for others, we made use of the Wason selection task in a similar way to Cosmides and Tooby (1992). An altruist-detection version of the Wason selection task needs to test whether people search for the presence of genuine emotions behind an altruistic act. The altruist-detection Wason task contains the following rule structure: "If X helps, then X seeks credit." Subjects then are asked who they would sooner trust. Specifically, subjects can choose any of the following cards: (1) "X helps"; (2) "X does not help"; (3) "X does not seek credit"; and (4) "X seeks credit." The correct answer is "X helps" and "X does not seek credit." This person is the altruist who is most likely to cooperate in the future. However, the person who helps only to receive credit (choosing the cards "X helps" and "X seeks credit") presumably is less likely to help in the future when it is not in his or her self-interest.

The present study was designed to test for the existence of an altruist-detection algorithm. Specifically, we tested an altruist-detection problem against control Wason selection problems (e.g., abstract version of task and a familiar Wason problem on school assignment of students). It was predicted that a higher frequency of subjects would solve the Wason tasks designed to detect altruists than the control problems.

1. Methods: experiment 1

1.1. Subjects and instruments

Sixty subjects from Introductory Psychology classes at Dalhousie University were recruited for three problem-solving tasks; participants received a 1% credit toward their class grade. One problem was the abstract Wason task of Griggs and Cox (1983). This abstract Wason task was used as a control to determine a baseline frequency of correct choices for our sample on the traditional Wason task. In previous research, 4% to 10% of subjects typically get the abstract version of the Wason selection task correct (Wason, 1983). The correct answer in logical terms is "P" and "not-Q" for an "If P, then Q" statement. In our problem of letters and numbers (A = P, 5 = Q, 8 = not-Q, and K = not-P), the logically correct response is "A" and "8."

The two altruist-detection problems follow this "if P then Q" structure. Problem 1 (Appendix 1) asked the subjects to imagine that they are in need of a trustworthy babysitter. A genuinely altruistic babysitter in this Wason problem is the person who volunteers to help sick children for the sake of helping rather than consciously seeking self-gain. Problem 1's rule is: "If they volunteer, then they seek credit." The four card choices are (1) seeks credit; (2) does not seek credit; (3) volunteers; and (4) does not volunteer.

Altruist-detection problem 2 (Appendix 2) has the same rule structure as problem 1; however, the situation is different. Specifically, problem 2 asks subjects to imagine that they need

to choose a friend who can help them adjust to a new city without taking advantage of them. A likely candidate is the person who gives blood but does not accept the payment for doing so. Problem 2's rule is: "If they give blood, then they accept payment." The four cards read (1) accepts payment; (2) does not accept payment; (3) gives blood; and (4) does not give blood. In both altruist-detection problems 1 and 2, the logically correct choice ("P" and "not-Q") corresponds to the detecting the altruist (person helps and does not seek gain).

The switched versions of problem 1 and 2 are identical to the standard versions except the "If...Then" statement was reversed: "If they seek credit, then they volunteer" and "If they accept payment, then they give blood." The purpose behind switching the problems was to make sure that subjects were detecting motives rather than simply reasoning more logically. If subjects were reasoning logically, they would choose "P" and "not-Q." However, if the subjects are detecting altruists, they would choose "not-P" and "Q" in the switched version. Specifically, this manipulation can test the hypothesis of whether facilitated logical reasoning is improving performance, or subjects simply are choosing the cards denoting a genuine altruist ("X helps" and "X does not seek credit").

For all five problems subjects were instructed to "select only the card(s) they definitely need to turn over to determine whether" (1) "the rule has been violated" (abstract version); (2) "whether any of your co-workers are potential friends" (altruist-detection problem 1 standard and switched versions); and (3) "whether a candidate is acceptable to you as a baby-sitter" (altruist-detection problem 2 standard and switched versions).

1.2. Procedure

Subjects were brought into the lab in groups of 15. Problem order was counterbalanced to avoid any effect of order of presentation. Each subject received three problems; the two altruist detection problems, either a standard or switched rule versions, and an abstract (control) problem. Half the sample ($n = 30$) received a standard version of altruist-detection problem 1 and switched version of altruist-detection problem 2. The other half of the sample ($n = 30$) received a switched version of altruist-detection problem 1 and a standard version of the altruist-detection problem 2. All subjects received the standard control problem (abstract version). Subjects were given instructions before they began to work on the problems. Instructions were limited so not to bias performance. Subjects were instructed that their participation would require them to complete three problem-solving tasks.

1.3. Data analyses

To compare the frequency correct on the altruist-detection tasks versus control tasks, we made use of McNemar's test. According to Siegel (1988), McNemar's test is a suitable nonparametric statistic for comparing frequencies in matched samples. McNemar's test is based on a Chi-square test for goodness of fit that compares the distribution of counts expected under the null hypothesis to the observed counts. However, unlike the Chi-square test for goodness of fit and the Chi-square test for independence, it does not have the assumption of independent observations. We have a between-subjects and within-subjects design. Therefore, we used McNemar's test for the within-subjects comparisons and regular Chi-square tests for

the between-subjects comparisons. To distinguish between McNemar's Chi-square and regular Chi-square tests, we used a "M" prefix denoting a McNemar's test (e.g., $M\chi^2$).

1.4. Results

The content differences between altruist-detection problems 1 and 2 did not affect performance. Fifteen of 30 selected the "P" and "not-Q" cards (i.e., detected the altruist) in the standard version of altruist-detection problem 1 compared to 17 of 30 who selected the "P" and "not-Q" cards in altruist-detection problem 2. The proportion of correct choices for the two altruist-detection problems with the standard rules did not differ significantly [Table 1A; $M\chi^2$ (1) = 0.70, $p > .05$]. For the switched version of the altruist-detection problem 1, 18 of 30 selected the "Q" and "not-P" cards (i.e., detected the altruist) compared to the 17 of 30 who selected "Q" and "not-P" cards (i.e., detected the altruist) in altruist-detection problem 2. The proportion of correct choices for the two altruist-detection problems with the switched rule did not differ significantly [$M\chi^2$ (1) = 0.20, $p > .05$]. Therefore, the results of the two problems were combined for further analyses.

As predicted, more subjects (32/60) detected the altruist by solving the altruist-detection problems correctly compared to the number of subjects (1/60) who correctly solved the control abstract Wason task [Table 1A; $M\chi^2$ (1) = 29.03, $p < .00001$, $n = 60$]. Performance on the abstract task was somewhat lower in our sample compared to previous studies using the abstract Wason task (around 10%). Therefore, we performed a Chi-square test for goodness of fit comparing the frequency correct on the altruist-detection task versus the base-line frequency usually found in previous studies using the abstract Wason task. Consistent with the hypothesis, significantly more subjects (53%) got the altruist-detection version correct than

Table 1
Percentage of subjects choosing "P and not-Q" and "Q and not-P" for standard rules (altruist-detection, cheater-detection, abstract control, and "school problem" control) versus percentage of subjects choosing "Q and not-P" and "P and not-Q" for switched rules (altruist-detection, cheater dection, and "school problem" control)

	Percentage choosing	
Problem Type	P and not-Q	Q and not-P
A) Experiment 1		
Standard versions (if P, then Q):		
Altruist-detection	53	0
Abstract control	1.7	0
Switched version (if Q, then P):		
Altruist-detection	1.7	58
B) Experiment 2		
Standard versions (if P, then Q):		
Altruist-detection	56.5	0
Cheater-detection	69.6	0
"School problem" control	34.8	0
Switched versions (if Q, then P):		
Altruist-detection	13	60.4
Cheater-detection	4.3	69.6
"School problem" control	26.1	0

expected if there were no altruist-detection abilities in our subjects [Table 1A; χ^2 (1) = 125.19, $p < .0001$, $n = 60$].

We also hypothesized that the selection of "P" and "not-Q" cards would decrease for switched altruist-detection rules. Recall that the justification for this hypothesis is that people are not solving the problems logically; rather, they are attempting to detect the genuine altruist. It was found that 32 subjects selected the "P" and "not-Q" cards when they received a standard rule, and 35 subjects selected "Q" and "not-P" when they received the switched version of the same rule. Specifically, this finding indicates that subjects preferred the cards "X helps" and "X does not seek credit" in both standard and switched altruist-detection problems, even though this answer is "logically" incorrect in switched versions. The proportion of "P" and "not-Q" choices was significantly higher for the altruist-detection problem with the standard rule compared to the altruist-detection problem with the switched rule [Table 1A; χ^2 (1) = 29.03, $p < .00001$, $n = 60$]. This result suggests that the wording of the altruist-detection problems is not simply allowing subjects to think more logically, but rather subjects appear to be detecting altruists. In fact, as seen in Table 1A, 58% of subjects chose "not-P" and "Q" cards in the switched altruist-detection problem compared to the 1.7% who chose "P" and "not-Q."

1.5. Discussion

The main hypothesis of Experiment 1 was that subjects would be able to solve problems designed to detect altruists better than the abstract Wason control problem. The results were consistent with this hypothesis. More subjects detected the altruist than would be expected if humans did not pay particular attention to information related to a conspecific's underlying altruism. Altruist-detection was not mediated by logical reasoning as subjects solved switched altruist-detection problems illogically, but correct from the perspective of detecting the altruist. Specifically, the logically correct answer would be "P" and "not-Q," but most subjects chose "not-P" and "Q" for switched problems. Put into altruist-detection terms, most subjects chose the cards "Does not seek credit" and "X helps" regardless of whether subjects received standard or switched rules. It is this combination of cards that reflects genuine altruism.

The results of Experiment 1 suggest that people are superior at reasoning about altruistic intentions compared to abstract problems in Wason selection tasks. Furthermore, this result is not due to altruist-detection tasks facilitating logical reasoning. These results parallel those of Cosmides (1989), who found that cheater-detection tasks were more easily solved than the abstract version and that this effect was not mediated by logical thinking. In a second experiment, we explored these findings further in two ways. In Experiment 1, we used a very abstract control problem with <2% of our subjects solving it correctly. Subjects typically perform better with familiar content (Cosmides and Tooby, 1992). Therefore, in Experiment 2, a control Wason task was included with concrete wording (i.e., not letters and numerals) that was matched for approximate word count to that of the altruist-detection problem. In Experiment 2, performance on the altruist-detection problem was compared to performance on the "school problem," a control Wason selection task taken from Cosmides (1989).

Experiment 2 also made comparisons between a cheater-detection version of the Wason task and our altruist-detection version. No a priori predictions were made regarding subjects'

performance on the altruist-detection problem versus the cheater-detection problem. Without a good evolutionary model, it is difficult to predict on which type of problem (altruist vs. cheater detection) subjects would exhibit better performance.

The cheater-detection version of the Wason selection task follows a social exchange structure: "If X receives the benefit, then X must pay the cost." One side of the card tells whether or not X takes the benefit, and the other side of the card tells the subject whether or not X paid the cost. The four cards presented are (1) "X takes benefit"; (2) "X does not pay the cost"; (3) "X does not take benefit"; and (4) "X pays the cost." Subjects are asked to determine whether any individual is violating this rule. The correct answer is "X takes benefit" and "X does not pay the cost." About 76% of subjects get this problem correct (Cosmides and Tooby, 1992). As before, the altruist-detection and the cheater-detection problems were switched. We also included a switched version of the school control problem. It was predicted that subjects would choose "P" and "not-Q" cards for standard altruist-detection and standard cheater-detection problems, but when the problem rules are switched we expected that subjects would detect altruists and cheaters by choosing "not-P" and "Q." However, when it comes to the school control problem, we expected that switched rules would not increase "not-P" and "Q" card choices.

2. Methods: experiment 2

2.1. Subjects and procedure

Forty-six subjects from Introductory Psychology classes were brought into the laboratory in groups of approximately 15 and administered three Wason problems in counterbalanced order of presentation (4 of the 6 possible orders were used 8 times, the other 2 possible orders were used 7 times): (1) the altruist-detection problem (Appendix 2); (2) the cheater-detection problem (Cosmides and Tooby, 1992); and (3) the school problem (Cosmides, 1989). Half the sample ($n = 23$) received standard versions and the other half of the sample ($n = 23$) received switched versions. Before subjects were permitted to attempt the problems, they were given the same brief instructions as in Experiment 1. They also were given class credit and recruited in the same manner.

The altruist-detection problem used in Experiment 2 was used in Experiment 1 (Appendix 2). Recall that a correct answer on an altruist-detection task consist of choosing the "X helps" and "X does not seek credit" cards. The cheater-detection problem is a standard social contract ("If X receives the benefit, then X must pay the cost"). The version used was the Drinking Age problem (Cosmides and Tooby, 1992) in which subjects are presented the conditional statement: "If a person is drinking beer, then they must be over 18 years old." In a cheater-detection problem, the correct answer is "benefit accepted" and the "cost not paid" cards. Since we are using the Drinking Age problem, the functional equivalent answers are the "drinking beer" and "16 year old" cards. The School Problem contains the conditional statement: "If a student is to be assigned to Halifax High School, then that student must live in Halifax." The card choices are Halifax High School, Dartmouth, Halifax, and Dartmouth High School. For the "School Problem" the logical correct answer is "P" (Halifax High

School) and "not-Q" (Dartmouth). Whether the "School Problem" rule is standard or switched, subjects should still choose "P" and "not-Q."

2.2. Results

Thirteen subjects selected "P" and "not-Q" cards for the altruist-detection problem with standard rules, whereas 16 and 8 selected "P" and "not-Q" cards for the cheater-detection and school control problems with standard rules, respectively. No statistically significant differences were found between cheater-detection and altruist-detection problems with standard rules [Table 1B; $M\chi^2$ (1) = 0.45, $p > .05$]. In addition, there was no statistically significant difference in performance on the altruist-detection problem versus the school control problem with standard rules [Table 1B; $M\chi^2$ (1) = 0.64, $p > .05$, n =23]. There was no significant difference between the frequency of subjects selecting the "not-P" and "Q" cards for altruist-detection and cheater-detection versions with switched rules [Table 1B; $M\chi^2$ (1) = 0.08, $p > .05$, n =23]. Finally, significantly more subjects selected "not-P" and "Q" when solving a switched altruist-detection (14/23) compared to the number of subjects who selected "P" and "not-Q" on the school control problem (6/23) [$M\chi^2$ (1) = 4.08, $p < .05$, n = 23].

Descriptive analyses in Table 1B revealed that for the altruist-detection and cheater-detection problems, subjects chose "not-P"s and "Q"s when the problems were switched. However, this trend was absent for the school control problem. As in Experiment 1, the switched altruist-detection problem had significantly fewer "P" and "not-Q" choices than the standard altruist-detection problem [Table 1B; χ^2 (1, n = 46) = 17.67, $p < .0001$]. Table 1B shows that 56.5% of subjects chose "P" and "not-Q" cards in standard problems (only 13% of subjects chose "P" and "not-Q" in switched problems). This result suggests that subjects are attempting to detect altruists rather than solving the problem via logical thinking.

An analogous effect appears for the cheater-detection problem. Specifically, as seen in Table 1B, 69.6% of subjects chose "P" and "not-Q" card when solving a standard cheater-detection task; however only 4.3% chose "P" and "not-Q" when the same problem rule is switched. This difference was significant [χ^2 (1, n = 46) = 46.28, $p < .0001$]. This is consistent with previous findings suggesting that subjects are detecting cheaters rather than simply solving the problem logically.

However, as expected, this "switching effect" disappears for the school control problem. Specifically, as seen in Table 1B, "P" and "not-Q" was chosen approximately the same number of times for standard (34.8%) as well as the switched rule (26.1%). This difference was not significant [χ^2 (1) = .77, $p > .38$, n = 46]. This suggests that subjects who do manage to solve the task chose the "P" and "not-Q" cards regardless of the switching of the conditional rule.

Given that the same patterns exists across standard and switched versions, the data for the different versions were combined for further analyses. When the frequency correct for standard and switched were combined, more subjects (27/46) were correct on the altruist-detection problem than on the school control problem (15/46) [Table 1B; $M\chi^2$ (1) = 6.86, $p < .01$, n = 46]. Likewise, the frequency of subjects solving cheater-detection problems correctly (32/46) was significantly greater than the frequency solving the school control problem correctly (15/46)[Table 1B; $M\chi^2$ (1) = 14.09, $p < .0001$, n = 46]. However, the frequency of

subjects (32/46) correctly identifying the cheater did not differ significantly from the frequency of subjects (27/46) correctly identifying the altruist [Table 1B; $M\chi^2$ (1) = .45, $p > .05$, $n = 46$].

2.3. Discussion

As in Experiment 1, subjects tended to perform better on altruist-detection tasks (i.e., detected the altruist) than on the control task. This result did not appear to be mediated by logical reasoning. This finding provides more support for the idea that humans scrutinize the proximate motives why an individual was altruistic. This is exactly what we expect if subtle cheating has exerted selection pressures on human cognitive architecture to search for information regarding the genuineness of an altruistic act.

The second finding in Experiment 2 was that there was no significant difference in performance on the altruist-detection versus the cheater-detection versions. It is difficult to draw definite conclusions from this null result. It is possible that the sample size was too small to detect a significant effect. However, it is clear that both cheater-detection and altruist-detection are occurring in these situations.

3. General discussion

This study suggests that participants can solve altruist-detection Wason problems significantly better than control Wason problems (abstract and school versions). This finding is analogous to the findings on cheater-detection by Cosmides (1989). Altruist-detection tasks elicit the switching response (i.e., "not-P" and "Q") for reversed rules. The switching response supports the notion that subjects are detecting altruists rather than giving logically correct responses. Furthermore, participants did not solve cheater-detection tasks better than altruist-detection tasks. These results are consistent with the evolutionary hypothesis that humans are sensitive to information regarding the genuineness of the altruistic behaviour performed.

The result that participants can detect altruists in a Wason selection task at first glance contradicts a finding by Cosmides and Tooby (1992) suggesting that subjects could not detect altruists. However, their "altruist version" Wason task was not designed with reference to any theory that predicts the existence of altruist-detection (Frank, 1988). Rather, the altruist version Wason task of Cosmides and Tooby (1992) was used as a control, and the character was not portrayed as a genuine emotion-based altruist, but as one who displayed occasional "random" acts of indulgence. It is possible that one of the reasons that subjects performed poorly on their "altruist version" is that their story context was not appropriate for testing the hypothesis of genuine altruism-detection proposed by Frank (1988).

Cosmides and Tooby (1992) have proposed the existence of cheater-detection information processing mechanisms. This is what a neo-Darwinian approach would predict in a species that routinely engages in social contracts where detection of cheating is necessary for protection from free-riders (Axelrod and Hamilton, 1981; Trivers, 1971). Experiments have generally presented convincing evidence of the existence of a cheater-detection information processing mechanism in humans (Mealey et al., 1996; Oda, 1997).

When subjects solve a Wason task correctly, can it be concluded that there exists an altruist-detection "Darwinian Algorithm"? Maybe subjects are good at reasoning about intentions

rather than altruist-detection per se. Likewise, humans could have refined abilities for detecting individuals to exploit. And genuine intentions could be a signal of who can be exploited, that is, who is a "sucker." However, as long as altruists' detection abilities are no worse than exploiters' detection abilities, they can form partnerships to avoid such exploitation.

An alternative interpretation of our findings is that our altruist-detection problems activated a social exchange "search for cheaters" algorithm. Specifically, participants may have read the problem as asking, "who is definitely not the cheater?" Future research needs to clarify whether an "altruist-detection cognitive mechanism" functions independently from a "cheater-detection cognitive mechanism." We feel that detecting cheaters in a social exchange situation is different from detecting altruistic commitments or aspects of an individual's character. That is, the work by Cosmides (1989) demonstrates that humans are adept at reasoning about cheating in social exchange situations, whereas the present study suggests that humans are good at reasoning about attributes indicating an altruistic character. Tooby and Cosmides (1996) also suggested that, evolutionarily, we should expect to find different psychological mechanisms in the domain of friendship as opposed to social exchange.

Consistent with the current study but using a different methodology, Brown (1998) found that subjects could detect altruists. Specifically, subjects could detect self-reported altruists after a 1-minute video-clip presentation. Subjects rated altruists as more concerned, attentive, and helpful than nonaltruists. Therefore, it could be a possibility that subjects who solve altruist-detection Wason tasks may be utilizing the same altruist-detection capacities as the subjects who detect altruists in video clips.

Evolutionary researchers have implied that cooperators may assortate with one another to gain the benefits of mutual cooperation (Bull and Rice, 1991; Peck 1993; Wilson and Dugatkin, 1997). One adaptive benefit of reliably detecting altruists is that an altruist would accrue inclusive fitness benefits when they gain allies and friends to rely on for hunting, gathering, and social activities (Alexander, 1987; Tooby and Cosmides, 1996). What is needed for these models is a theory for the evolution and stability of altruists reliably signaling their intentions (Grafen, 1990; Zahavi, 1987). Perhaps deceptive displays of pseudo-altruism were kept in evolutionary check via frequency dependent selection (Frank, 1988; Trivers, 1985; see also Surbey and McNally, 1997 for a study of the psychological mechanisms involved in this proposed coevolutionary struggle). Semple and McComb (1996) have suggested that selection may favor honesty when the high costs of being deceived leads to the selection of skeptical perceivers who only respond to intrinsically unfalsifiable signals. Frank (1988) has offered such a mechanism, and evidence consistent with one of the assumptions of his model (i.e., altruist-detection) has been provided here.

Acknowledgments

We thank Heather Munro for collecting the data in Experiment 1 as part of her undergraduate honor's thesis. For useful criticisms of the manuscript, we thank Margo Wilson and Martin Daly. We also would like to thank two anonymous referees for suggesting the alternate explanations for our findings. This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Postgraduate Scholarship and the Isaak

Walton Killam Memorial Scholarship to William Michael Brown and by Social Science and Humanities Research Council of Canada grant 410-98-0462 to Chris Moore.

Appendix I

Altruist-Detection Wason Selection Task–Version 1

Imagine that you have had a newborn baby within the past year and you decided to go back to work. You are now in need of a trustworthy babysitter. Because there has been recent media reports of baby-sitters who have abused children, you have to be extra careful to select a sitter who will genuinely care for your child. But you do not want to hire someone simply because they have babysat before. Instead, you wish to base your decision on how genuinely concerned the person is for the welfare of others. This quality often can be demonstrated when people volunteer within the community without receiving material rewards of any kind.

Therefore, you decide to hire someone who volunteers to help sick children on his or her days off for the sake of helping rather than for self-gain or academic credit (for example: volunteering for extra school credit or volunteering just to improve his or her resume).

As a result, those candidates who observe the following rule are considered unacceptable to care for child:

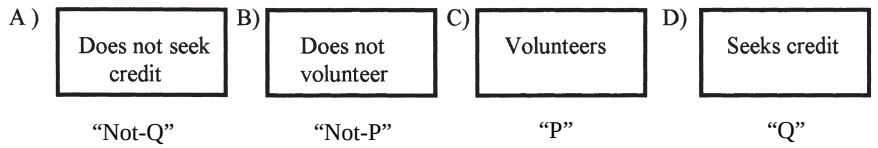
"If they volunteer, then they seek credit." (standard version)

OR

"If they seek credit, then they volunteer." (switched version)

The cards below have information about four potential candidates. One side of each card tells you whether or not a candidate volunteers and the other side of each card tells you whether or not they seek credit.

Choose only the card(s) you definitely need to turn over to determine whether a candidate is acceptable to you as a babysitter.



Appendix 2

Altruist-Detection Wason Selection Task–Version 2

You have been offered an excellent job in your field in New York City. Although you are excited about this great career opportunity, you are worried about finding friends who can help you deal with the transitional adjustments of living in a new city. Coupled with the fact that recent studies have shown that there are many people who cannot be trusted in New York City, your worries are justified.

You would like to have close friends who will not take advantage of you in the workplace, nor in personal life. You wish to base your friend choice on how genuinely concerned they are for others. Thus, you decide to befriend anyone who gives to others and does not ask for anything in return.

In the same building where you work, a health clinic has set up temporary facilities for giving blood. Many people from your office plan to give blood, and you consider this a good opportunity to meet potential friends.

The clinic is desperately in need of blood supplies and is willing to offer a small cash payment to each person who gives their blood. Of course, the idea of accepting payment for such a good deed is not something you would do. Similarly, you consider anyone who does accept payment for giving blood to not be as selfless as they appear, and thus not someone you wish to befriend.

Therefore, those co-workers who follow the rule below are not considered to be an accepted friend to you:

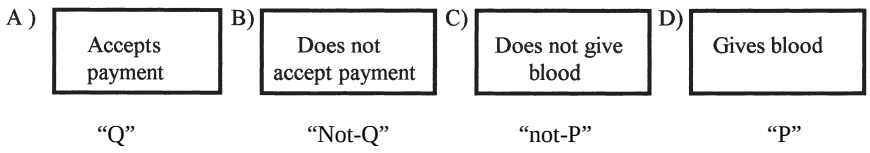
"If they give blood, then they accept payment." (Standard version)

OR

"If they accept payment, then they give blood." (Switched version)

The four cards below have information about four co-workers. One side of each card has information about whether or not they gave blood, and the other side of each card has information about whether or not they accepted payment.

Indicate only the card(s) that you definitely need to turn over to determine if any of your co-workers are potential friends.



References

Alexander, R. D. (1987). *The Biology of Moral Systems*. New York: Aldine de Gruyter., 1987.

Axelrod, R. & Hamilton, W. D. (1981). The evolution of cooperation. *Science* 211, 1390–1396.

Brown, W. M. (1998). *Signaling Benevolence: An Evolutionary Perspective on the Encoding and Decoding of Altruism.* Master of Science Thesis, Department of Psychology, Dalhousie University: The National Library of Canada.

Bull, J. J. & Rice, W. R. (1991). Distinguishing mechanisms for the evolution of cooperation. *Journal of Theoretical Biology* 149, 63–64.

Cooper, B. & Wallace, C. (1998). Evolution, partnerships and cooperation. *Journal of Theoretical Biology* 195, 315–329.

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31, 187–276.

Cosmides, L. & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 163–228). Oxford: Oxford University Press.

Dugatkin, L. A. (1997). *The Evolution of Cooperation*. Oxford: Oxford University.
Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions.* New York: W.W. Norton and Co.
Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: an experiment. *Ethology and Sociobiology* 14, 247–256.
Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology* 144, 517–546.
Griggs, R. A. & Cox, J. R. (1983). The effects of problem content and negation on Wason's selection task. *Quarterly Journal of Experimental Psychology* 35A, 519–533.
Hamilton, W. D. (1971). Selection of selfish and altruistic behaviour in some extreme models. In J. F. Eisenberg and W. S. Dillon (Eds.). *Man and Beast: Comparative Social Behavior* (pp. 57–91). Washington, DC: Smithsonian Press.
Hirshleifer, J. (1987). On the emotions as guarantors of threats and promises. In J. Dupre (Ed.). *The Latest on the Best: Essays on Evolution and Optimality* (pp. 307–326). Cambridge: MIT Press.
McCabe, K. A., Rassenti, S. J., & Smith, V. L. (1996). Game theory and reciprocity in some extensive form experimental games. *Proceedings of the National Academy of Sciences* 93, 13421–13428.
Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology* 17, 119–128.
Nowak, M. & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* 364, 56–58.
Oda, R. (1997). Biased face recognition in the prisoner's dilemma game. *Evolution and Human Behavior* 18, 309–315.
Palameta, B. & Brown, W. M. (1999). Human cooperation is more than by-product mutualism. *Animal Behaviour* 57, F1-F3 (http://www.academicpress.com/www/journal/ar/forum.htm).
Peck, J. R. (1993). Friendship and the evolution of cooperation. *Journal of Theoretical Biology* 162, 195–228.
Semple, S. & McComb, K. (1996). Behavioural deception. *Trends in Ecology and Evolution* 11, 434–37.
Siegel, S. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
Surbey, M. K. & McNally, J. J. (1997). Self-deception as a mediator of cooperation and defection in varying social contexts described in the Prisoner's Dilemma. *Evolution and Human Behavior* 18, 417–435.
Tooby, J. & Cosmides, L. (1996). Friendship and the banker's paradox: other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy* 88, 119–143.
Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology* 46, 35–57.
Trivers, R. L. (1985). *Social Evolution*. Menlo Park: Benjamin/Cummings.
Wason, P. C. (1983). Realism and rationality in a selection task. In J. St. B. T. Evans (Ed.). *Thinking and Reasoning: Psychological Approaches* (pp. 47–75). London: Routledge & Kegan Paul.
Wilson, D. S. & Dugatkin, L. A. (1997). Group selection and assortative interactions. *American Naturalist* 149, 336–351.
Zahavi, A. (1987). The theory of signal selection and some of its implications. In V. P. Delfino (Ed.). *Proceedings of the International Symposium of Biological Evolution* (pp. 305–327). Bari: Adriatica.