

Eye spots do not increase altruism in children*

Sonja Vogt^{1,2}, Charles Efferson^{1,2}, Joël Berger³, and Ernst Fehr^{1,2}

¹Department of Economics, University of Zurich

²Laboratory for Social and Neural Systems Research, University of Zurich

³Department of Humanities, Social and Political Sciences, Swiss Federal
Institute of Technology, Zurich

October 29, 2015

Running title: Eye spots do not increase altruism in children

Word count: c. 7000

***Corresponding authors:** Sonja Vogt (sonja.vogt@econ.uzh.ch), Charles Efferson (charles.efferson@econ.uzh.ch), and Ernst Fehr (ernst.fehr@econ.uzh.ch), Department of Economics, University of Zurich, Blümlisalpstrasse 10, 8006 Zurich, Switzerland.

Abstract: The evolutionary legacy hypothesis proposes that an evolved reciprocity-based psychology affects human behavior in anonymous one-shot interactions when reciprocity is not explicitly possible. Empirical support rests on experiments showing that altruism among adults increases in the presence of stylized eye spots or faces. Such stimuli do not affect material payoffs, but they are assumed to activate a person's reciprocity-based psychology. We identify two versions of the evolutionary legacy hypothesis. The weak hypothesis posits that reputational concerns can generate altruism in the absence of opportunities for a good reputation. The strong hypothesis posits that reputational concerns alone can explain anonymous one-shot altruism, and they can do so specifically in lieu of a explanations based on group selection. A number of experimental studies support the weak hypothesis but are merely consistent with the strong hypothesis. To address both the weak and strong hypotheses, we conducted an eye spot experiment with children. Altruism can vary by age or sex in childhood, and under the strong hypothesis this kind of variation should reveal associated variation in sensitivity to eye spots. Although we found significant variation in altruism among children, we found no corresponding variation in sensitivity to eye spots. More generally, we found no eye spot effects of any kind. We discuss the possibility that eye spots might only affect altruism under specific conditions. We further argue that conditional effects do not refute the weak hypothesis in any way, but they do suggest potential limitations on the explanatory scope of the strong hypothesis.

Key words: altruism, evolutionary legacy, dictator game, social preferences in children

1 Introduction

Altruistic cooperation and the prosocial preferences on which it depends play a crucial role in human societies (Bowles and Gintis, 2011). When social interactions take place in their usual setting, a setting characterized by incomplete contracts, social norms, and informal institutions, prosocial behavior can dramatically attenuate the inefficiencies that follow from strictly self-regarding behavior (Bowles, 2004). Nonetheless, in spite of the crucial role prosocial behavior has in human societies, the evolutionary mechanisms responsible for such behavior remain a highly contentious matter. This is especially true in the special and important case of altruism in anonymous one-shot interactions with genetically unrelated partners (Henrich, 2004; Burnham and Johnson, 2005; Haley and Fessler, 2005; Hagen and Hammerstein, 2006; Burnham, 2013).

Anonymous one-shot interactions are special because ethnographic data suggest that people rarely had anonymous interactions in ancestral societies (Fehr and Henrich, 2003). Moreover, some researchers additionally argue that ancestral social interactions were typically repeated. If so, one-shot interactions were probably also quite rare (Burnham and Johnson, 2005; Haley and Fessler, 2005; Hagen and Hammerstein, 2006). Although the nature of social life in the distant past will always involve some speculation, the evidence overall suggests that anonymous one-shot interactions constitute a special class of evolutionarily recent phenomena.

Anonymous one-shot interactions are additionally important for the following reasons. First, much of the experimental research documenting human altruism is based on interactions of this sort (Kagel and Roth, 1995; Camerer, 2003; Henrich *et al.*, 2004, 2006, 2010a). Second, even if nameless and ephemeral interactions were rare in the past, they are presumably quite common now, and for this reason they matter in contemporary human societies. Finally, the evolutionary basis for anonymous one-shot altruism among unrelated strangers is especially hard to identify and explain. In particular, the only evolutionary explanations for prosocial behavior that are widely regarded as unproblematic are kin-based altruism and the enlightened material self-interest of reciprocity in its various forms (Hamilton, 1964; Trivers, 1971; Axelrod and Hamilton, 1981; Nowak and Sigmund, 1998;

Henrich, 2004; Panchanathan and Boyd, 2004; Bowles and Gintis, 2011). With respect to anonymous one-shot altruism, however, neither kinship nor reciprocity provides an obvious explanation. The apparent alternative is group selection, especially selection between groups with different culturally transmitted social norms (Fehr and Fischbacher, 2004; Henrich, 2004; Richerson and Boyd, 2005; Bowles and Gintis, 2011). An active and persistent debate, however, has surrounded the plausibility of cultural group selection since the idea was first proposed as part of the more general research program on gene-culture coevolution (Boyd and Richerson, 1985, 2005; Burnham and Johnson, 2005; Haley and Fessler, 2005; Richerson and Boyd, 2005; Lehmann *et al.*, 2008; Bell *et al.*, 2009; Bowles and Gintis, 2011; Boyd *et al.*, 2011).

A clever and influential approach to this seeming impasse is to transfer unproblematic explanations based on reciprocity to situations in which the explicit material structure of a social interaction does not allow for reciprocity. Doing so is feasible insofar as the implicit structure of the interaction does allow for reciprocity. This is the evolutionary legacy hypothesis for the evolution of human altruism (Burnham and Johnson, 2005; Haley and Fessler, 2005; Hagen and Hammerstein, 2006; Burnham and Hare, 2007; Burnham, 2013). The hypothesis posits that human psychology reflects ancestral conditions that differed radically from the anonymous one-shot conditions implemented in many contemporary behavioral experiments.

Specifically, the hypothesis proposes that for much of our evolutionary past human social groups were small and cohesive, social contact was intense, interactions were repeated, and one's reputation as a more or less cooperative individual was always at stake. Life was effectively like a "camping trip that lasted a lifetime" (Cosmides and Tooby, 2013, p. 203). Under circumstances of this sort, a person would have cooperated in accord with reciprocal strategies that protected her reputation as a prosocial individual, and she would have done so to gain the benefits that followed when members of her group reciprocated in the future. Contemporary behavior stems from a psychology adapted to these ancestral conditions. Consequently, the behavior observed in contemporary anonymous one-shot interactions does not respond fully to the anonymity and transience of the setting. Put differently, even if the explicit structure of the interaction is anonymous and

one-shot, the implicit structure in the mind of the actor is such that she will behave in a way that bolsters her good reputation. In this sense, the evolutionary legacy hypothesis is fully consistent with evolutionary psychology more generally and its key principle that human cognition consists of modular adaptations to life as hunter-gatherers under ancestral conditions (Smith, 2000; Haley and Fessler, 2005; Kurzban, 2010; Burnham, 2013; Cosmides and Tooby, 2013).

The empirical evidence for a reputational psychology in anonymous one-shot interactions is typically experimental. In particular, a number of studies have shown that exposure to face-like stimuli increases altruistic choices in incentivized economic games (reviewed in Nettle *et al.*, 2013; Sparks and Barclay, 2013). Crucially, this outcome occurs even though the face-like stimuli are sometimes quite abstract and do not affect material payoffs. The resulting conclusion is that face-like stimuli, relative to control stimuli, activate the ancestral, reputation-based psychology of participants, and this leads to the observed increase in altruism. Simply put, if people feel they are being watched, they will behave themselves because future benefits from others depend on it.

These experimental findings are consistent with two basic versions of the evolutionary legacy hypothesis. The weak hypothesis is simply that reputational concerns can affect behavior in anonymous one-shot settings. We call this version “weak” because it does not exclude the possibility that other forces, even a group-selected psychology, are also at work. The “strong” hypothesis, however, does exclude other forces. It specifically posits that *every* anonymous one-shot interaction involves a variety of uncontrolled cues (e.g. other people in the room) that imply one’s reputation is at stake, and these cues produce *all* of the ostensibly other-regarding behavior observed in anonymous one-shot interactions. This possibility is especially important in terms of drawing inferences about evolutionary mechanisms from anonymous one-shot behavioral experiments among genetically unrelated strangers (Camerer, 2003; Henrich *et al.*, 2010b).

Importantly, we do not know of researchers previously using the terms weak and strong to discuss different versions of the evolutionary legacy hypothesis. Nonetheless, the terms capture different themes in the existing literature. The numerous studies showing that payoff-irrelevant faces increase altruism (reviewed in Nettle *et al.*, 2013; Sparks and Barclay, 2013) support the weak hypothesis.

They demonstrate that we cannot categorically ignore reputation and reciprocity simply because the explicit material structure of a game is anonymous and one-shot. Modulating altruism with faces, however, neither implies nor precludes effects associated with other evolutionary forces. For this reason, without additional arguments, existing experimental studies support the weak hypothesis. They are consistent with the strong hypothesis, but they do not provide direct support.

Nonetheless, some researchers have argued that we should grant reputation a kind of privileged explanatory status specifically when considering the evolutionary origins of anonymous one-shot altruism among genetically unrelated strangers (Burnham and Johnson, 2005; Haley and Fessler, 2005). The reason is straightforward. An evoked reputational psychology depends on an evolutionary history involving some kind of repeated interactions and the reciprocal strategies they support. The relevant evolutionary mechanisms operate via the long-term self-interest of individual organisms, and in this sense they are conventional, well understood, and uncontroversial (Burnham and Johnson, 2005; Burnham, 2013). In contrast, group selection, both cultural and genetic (Henrich, 2004; Bowles, 2006, 2009), represents a class of evolutionary mechanisms surrounded by controversy for nearly 50 years (Williams, 1966). Consequently, if we have no conclusive reason to reject reputation as an insufficient explanation, we should favor reputation in lieu of group selection (Burnham and Johnson, 2005; Haley and Fessler, 2005). More to the point, if we know payoff-irrelevant social stimuli matter, and if we know we can never eliminate such stimuli entirely, then reputation is all we need to explain altruism among anonymous unrelated strangers engaged in one-shot exchange. We do not need some form of group selection; nor should we turn to some form of group selection. When taken to its logical extreme, this argument leads to the strong version of the evolutionary legacy hypothesis. To be precise, we see the strong hypothesis as primarily relevant for understanding anonymous one-shot altruism among genetically unrelated strangers. For interactions of this sort, two basic evolutionary mechanisms are on the table: implicit reputation arising from an evolutionary history of repeated interactions and group selection. If genetic and cultural group selection are eliminated, we are left with the strong hypothesis.

The evolutionary legacy hypothesis has inspired a number of studies, and

many of them have provided compelling empirical support for the effects of payoff-irrelevant social cues. Importantly, however, previous research leaves us with two challenges. First, as discussed above, existing evidence does not allow us to evaluate the strong hypothesis. Assessing the strong hypothesis requires an approach with predictions that go beyond saying that reputational cues should increase altruism. Second, though many studies have provided evidence for the importance of payoff-irrelevant social cues (Haley and Fessler, 2005; Bateson *et al.*, 2006; Burnham and Hare, 2007; Rigdon *et al.*, 2009; Mifune *et al.*, 2010; Ernest-Jones *et al.*, 2011; Oda *et al.*, 2011; Francey and Bergmüller, 2012), the evidence overall is mixed. Some studies have also failed to find an effect (Fehr and Schneider, 2010; Lamba and Mace, 2010; Carbon and Hesslinger, 2011; Tane and Takezawa, 2011; Ekström, 2012; Raihani and Bshary, 2012).

To address these challenges, we conducted a dictator game study with payoff-irrelevant stimuli and a distinctive but informative subject pool. Like previous studies, our experimental treatments involved either an asocial control stimulus or a social, face-like stimulus. Unlike previous studies, however, we recruited children of ages five and eight of both sexes to participate in our study. Children represent an informative subject pool because previous research has shown that young children, like adults, care about their reputations (Engelmann *et al.*, 2012, 2013), and thus they have a reputational psychology that can be experimentally manipulated. Moreover, past research also suggests that altruistic behavior changes between the ages of five and eight, and this developmental process is potentially different for girls and boys (Benenson *et al.*, 2007; Fehr *et al.*, 2008; Blake and McAuliffe, 2011; House *et al.*, 2013). For these reasons, children provide a potential means of addressing both the weak and strong hypotheses.

Recall that the strong version of the evolutionary legacy hypothesis posits that all anonymous one-shot altruism among unrelated strangers is due to an active reputational psychology. By extension, under the strong hypothesis, which puts aside all other factors that can vary among people, any group of people that is relatively altruistic in an anonymous one-shot setting can only be this way because group members are relatively sensitive to the payoff-irrelevant social cues in their environment. For our experiment the strong hypothesis thus implies the following. If any group of children, defined by age, sex, or both, is systematically

more altruistic than other groups of children, this must be because the children in the altruistic group are more sensitive to the payoff-irrelevant social stimuli in our experiment. Such stimuli are the only means of modulating the reputational psychology that, under the strong hypothesis, is exclusively responsible for anonymous one-shot altruism. This reasoning applies to both the stimuli under experimental control and the stimuli not under experimental control. With respect to stimuli under experimental control, greater sensitivity by a specific group of children implies a positive interaction between being a member of the relatively altruistic group, whatever it may be, and exposure to our face-like stimulus. Because our design varies payoff-irrelevant stimuli for both girls and boys of different ages, it allows us to test for exactly this kind of interaction. As such, it provides a novel approach to evaluating the strong version of the evolutionary legacy hypothesis. More broadly, a pure treatment effect would support the weak hypothesis. Additionally, if a specific group of children is significantly more altruistic than other groups, a positive interaction between membership in this group and participation in our face treatment would support the strong hypothesis.

2 Experimental Methods

We conducted our experiment in the spring and fall of 2012 in 27 kindergarten and primary schools in the German-speaking part of Switzerland. The study was approved by all relevant local educational authorities and by the Human Subjects Committee of the Faculty of Economics, Business Administration, and Information Technology at the University of Zurich. Finally, we received informed consent from the parents of all participating children.

After arrival, while the children were in another area with their teacher, at least two experimenters constructed four arenas for the experimental session. These arenas consisted of two waiting areas, an area for the dictator game itself, and a toy store. After the experimental arenas were built, the experimenters brought the children together and introduced themselves. The experimenters explained that they came from the University of Zurich, and they were going to let each child play a game for the next hour or so. During the game, the child would have the opportunity to earn gold coins based on her choices in the game. In actuality,

the gold coins were simply stylized gold coins printed on small pieces of paper, and the experimenter took great care to explain the meaning and value of these gold coins.

Specifically, one of the experimenters held up a single gold coin and explained their value. At the end of the session, each child would have a chance to turn in her gold coins for a toy from the toy store. More coins would provide a child with the option to choose a toy from a larger selection that included larger and more valuable toys. At this point, the experimenter led the children to the toy store, where they observed different groups of toys separated both by value and whether they were targeted at girls or boys. After the children saw the toy store, the experimenters covered all toys with a blanket so the children would not be distracted during the experiment itself.

After this brief introduction, the experimenters led participants to the first waiting area. As mentioned above, we used two waiting areas, and we always placed them as far apart as possible. One waiting area was for participants waiting to play the game, while the other was for participants who had already played. Children solved puzzles in both waiting areas, and together the experimenters and the class teacher ensured that each child was in the appropriate waiting area. Importantly, the use of two waiting areas allowed us to separate experienced participants from naive participants. This in turn meant that the children who had played could not distort the play of the children still waiting to play by providing uncontrolled information about the game or recommendations about what to choose.

A single experimenter had primary contact with all participants. We did not inform the experimenter about our hypotheses, but we cannot categorically exclude the possibility that he drew his own inferences in a way that might have affected his behavior. In any case, we developed a protocol that was standardized across treatments (electronic supplementary material), and we trained the experimenter to follow this protocol assiduously in order to avoid cueing participants in any way. In addition, participants were randomly assigned to treatment within each session, and so we can also exclude the possibility of some unknown systematic relationship between session and treatment. For the dictator game itself, the principal experimenter led a randomly selected child to a small tent that was con-

structed away from the waiting areas. The child’s teacher was not present at this point. The child sat inside the tent, while the experimenter sat in the doorway of the tent. The child sat in front of a laminated pad approximately 30×42 cm. The lower half of the pad was blank, and 10 gold coins were placed on this blank area. The upper half of the pad consisted entirely of a graphical figure. In the control treatment, the figure was a meaningless but regular set of perpendicular lines (electronic supplementary material). In the eye spot treatment, the figure was the same stylized eye spots used in Haley and Fessler (2005).

To one side of the pad was a sealed box with a small opening, and to the other side of the pad were two envelopes. We counterbalanced the spatial orientation of the child and the experimental materials across sessions (electronic supplementary material, Fig. S3). One of the two envelopes was white, and it was for the participant’s own coins. The other envelope was one of four colors, and it was for the recipient’s coins. We used the colored envelopes to record the gender and treatment for each participant while still maintaining the participants’s complete anonymity. Specifically, the recipient’s envelope was either red, blue, green, or yellow. For each participant, the experimenter used an envelope of a specific color to record the appropriate gender-treatment combination for that participant. To avoid systematic effects associated with color preferences among the children, the meaning of the four colors was randomly determined for each session. As explained below, the child left the colored envelope behind after making her decision and leaving the tent. The color of this envelope allowed us to later identify the gender and treatment for the participant without the need to assign an identification card, ask the participant her name, or take any other action that would reduce the explicit anonymity of the child.

The experimenter explained the dictator game as follows. Specifically, the experimenter explained that the dictator’s task was to divide the 10 gold coins between herself and a recipient. At the end of the session, the dictator would have the opportunity to turn in her gold coins for a toy from the toy store, and more gold coins would allow her to choose from a larger set of toys that included more valuable toys. The recipient was another child in a different kindergarten or primary school, and neither the identity of the dictator nor the identity of the recipient would ever be revealed to the other. The experimenter explained

that the recipient had no coins herself. Any coins received from the dictator, however, would be turned in for a toy, and recipients with more gold coins would get more valuable toys. The dictator was told that, when making her decision, she should place her coins in the white envelope and the recipient's coins in the colored envelope. The experimenter explained that the dictator would be alone in the tent at this point, and the dictator should wear a nearby set of sound-proof earmuffs (Haley and Fessler, 2005). These earmuffs were especially designed for children, and our intention was to use both the tent and the earmuffs to eliminate uncontrolled social stimuli and so isolate the effect of our treatment variation. After distributing the coins, the experimenter told the dictator that she should deposit the recipient's envelope in the sealed box in the tent and then exit with her own envelope. In addition, the experimenter encouraged the dictator to put her envelope in her pocket, if possible, before proceeding to the second waiting area. Altogether, our experimental design achieved a degree of anonymity well beyond other studies conducted with children of this young age (cf. Fehr *et al.*, 2008).

Before leaving the child alone to make a choice, the experimenter asked the child several control questions to test understanding (electronic supplementary material). If the child answered any question incorrectly, the experimenter explained the task again and only continued when the child answered all questions correctly. Once the child had answered all control questions correctly, the experimenter helped the child put on her sound-proof earmuffs, closed the tent door so that the tent was opaque from all directions, and left the child to make her choice in private.

Altogether, we conducted the experiment with 201 children. 24 of them left the tent without dividing the coins, which leaves a total of 177 observations (94 boys, 83 girls). If the participant distributed the coins but forgot to take her envelope when exiting the tent, we recorded this fact and control for it in regression analyses below. This occurrence was rare and only happened with 12 dictators. When it did happen, the experimenter gave the child her envelope immediately. If the participant distributed the coins but forgot to deposit the recipient's envelope in the sealed box, we also recorded this information and control for it in our analyses. This happened 44 times. Of these 44 dictators, six of them were also

among the 12 dictators who forgot to take their own envelopes¹. After a particular participant was finished, the experimenter ensured that the participant went to the appropriate waiting area, and then the experimenter returned to the tent to prepare the task for the next participant.

At the end of each session, dictators went to the toy store one by one. Each dictator then turned in her gold coins for the toy of her choice from the toys she could afford with the number of coins she had. For each recipient, we translated the coins received into a toy from the appropriate price category (electronic supplementary material, Table S1). We then delivered the toys to a Swiss non-governmental organization that had previously agreed to distribute the toys to specific children of the appropriate ages from families in need.

Importantly, children often returned to their class with their toys after the experiment. Because toys were observable in such cases, they could have potentially affected a child's explicit reputation in the group. Although we could not eliminate this possibility entirely, we took several steps to minimize explicit reputational effects because our task instead was to focus on implicit or evoked reputation. Most importantly, we always ran two entirely different experiments in a given session, and this necessarily prevented children from giving a common meaning to a given toy. Indeed, in all sessions many of the children present participated in an experiment involving a multi-period coordination game. Unlike the dictator game described in the present paper, payoffs in this coordination game experiment depended primarily on coordinating repeatedly.

Children were randomly divided into two pools for the two different experiments at the beginning of each session. They were told the tasks would be different, but they received no information about the experiment they did not do. Moreover, all children, regardless of which experiment they participated in, were paid from the same toy store. A number of sessions also included a third group of children who were not allowed to participate in an experiment but were allowed to receive a toy to avoid being excluded. These children typically worked on puzzles in the same waiting areas used by children assigned to play the dictator game. Our pro-

¹Our overall impression was that many of the children who did not precisely follow all instructions regarding the envelopes failed to do so because they were very focused on the distribution task and the toy they would soon receive. As a result, when left alone without further reinforcement, they either did not see a reason to precisely follow all instructions regarding the envelopes, or they simply forgot to follow all instructions.

cedures ensured that every session involved two to three entirely different ways of receiving toys as payoffs, and our procedures made this clear in all sessions. As a result, children should not have had a common, consistent, and transparent basis for using toys to infer behaviors. Similarly, the class teacher also did not have the information to draw behavioral inferences. We did not explain either the coordination game or the dictator game to the teacher in detail. In addition, the children did not observe us informing the teacher, even in general terms, about the rules of either game.

3 Results

The mean transfer over all participants was 2.418, which is significantly positive (Wilcoxon signed rank test with continuity correction, $N = 177$, $p < 0.001$). Children randomly assigned to the control transferred an average of 2.552 gold coins, while children assigned to the eye spot treatment transferred an average of 2.289 gold coins. The difference is not significant (Wilcoxon signed rank test with continuity correction, $N = 177$, $p = 0.5152$). More generally, the transfer distributions are extremely similar for the treatment and control (Fig. 1a).

If we dichotomize transfers as zero or positive (Nettle *et al.*, 2013), 64.37% of participants in the control transferred some positive amount, while in the eye spot treatment 62.22% transferred a positive amount. Again, the difference is not significant (Fisher's exact test, $p = 0.876$). These results show that we have no pure treatment effect. Moreover, this is true even though we took great care to randomize assignment to treatment within sessions, to randomize the meaning of our colored envelopes in each session, to prevent any contamination of participants waiting to play, and to reduce uncontrolled social stimuli as much as possible in order to isolate the effect of our treatment variation.

Although we find no treatment effect over all participants, treatment effects could nevertheless be present but vary by the sex and age of the dictator. For example, perhaps only dictators in a particular age-sex category respond to eye spots. When all dictators are pooled, the effect becomes extremely difficult to detect. As an even more complex possibility, perhaps one age-sex combination increases transfers when assigned to the eye spot treatment, while another age-sex

combination decreases transfers. We conducted our experiment with girls and boys of ages five and eight to evaluate exactly this kind of possibility. More to the point, existing behavioral evidence suggests that prosocial behavior varies for children of ages five and eight, and the degree of prosocial behavior could also be different for boys and girls (Benenson *et al.*, 2007; Fehr *et al.*, 2008; Blake and McAuliffe, 2011; House *et al.*, 2013). If this is true, insofar as all prosocial behavior depends on payoff-irrelevant social cues, children in different age-sex categories should respond differently to these cues.

To evaluate this idea, we modeled transfers in two ways. First, we modeled player transfers using ordinal logistic regressions. Ordinal logistic regressions assume that the response variable falls into a relatively small number of discrete, ordered categories (Verbeek, 2008). In our case, ordinal logistic models are especially appropriate both because they do not assume that transfers represent a cardinal measure of preferences and because we have a number of observations at one of the boundaries of the action space (i.e. 0). Second, we also coded transfers as either zero (0) or positive (1) and modeled them using simple logistic regressions. This binary treatment of transfers allows us to identify variation in the probability of transferring any positive amount without regard for the magnitude of positive transfers (Haley and Fessler, 2005; Nettle *et al.*, 2013).

Under both treatments of the response variable, we specified six regression models. The models vary in terms of the independent variables they include (electronic supplementary material, Tables S2 and S3). All models control for whether the dictator dropped the recipient’s envelope in the sealed box in the tent and whether the dictator exited the tent with or without her own envelope. The set of independent variables additionally includes the age, sex, and treatment for each dictator. Altogether we specified models without interaction terms as well as models that include second-order and third-order interactions. Our complete set of models allows us to identify, with complete flexibility, any differential response to the eye spots based on the age of the dictator, the sex of the dictator, or both. Furthermore, this is true for both the regressions that model the entire space of dictator transfers (i.e. ordinal logistic) and the regressions that simply treat transfers as zero or positive (i.e. simple logistic).

To identify the best fitting models, we used AIC_c as a criterion to identify

the best models in the set (Burnham and Anderson, 2002). AIC_c is an information theoretic model selection criterion that optimizes the trade-off between the omitted-variable biases associated with underfitting and the lack of generalizable conclusions associated with overfitting. It is a derivative form of Akaike’s original asymptotic criterion (Akaike, 1973) that corrects for finite samples (Burnham and Anderson, 2002). We implemented the model selection exercise twice, once using ordinal logistic regressions and once using simple logistic regressions.

Both model selection exercises produced a robust result. Namely, the critical independent variable is the sex of the dictator (Fig. 1b). Boys transferred significantly less than girls (Table 1), and they transferred positive amounts with a significantly lower probability than girls (Table 2). Neither model selection nor associated regression results indicate that experimental treatment is an important predictor of dictator behavior. For both ordinal logistic regressions and simple logistic regressions, the best set of predictor variables includes the age and sex of the dictator, but no interactions and no variable for experimental treatment. Under this specification, whether using an ordinal logistic or a simple logistic regression, the sex of the dictator is highly significant, and it is the only significant predictor (electronic supplementary material). Here we do not present these best-fitting models simply because they do not include experimental treatment as a predictor. For the sake of generality, we instead present models that are similar to the best-fitting models but additionally include a treatment dummy (Tables 1 and 2). With or without the treatment dummy, the conclusions are the same. Boys gave less than girls, and they gave positive amounts with lower probability than girls.

4 Discussion and Conclusion

The evolutionary legacy hypothesis provides an influential explanation for anonymous one-shot altruism (Burnham and Johnson, 2005; Burnham, 2013). The hypothesis claims that altruism of this sort, under an appropriate evolutionary model of human psychology, does not need to be anonymous or one-shot. It can arise instead from an evolutionary history of reciprocity and reputation. Support for the hypothesis largely rests on experiments showing that payoff-irrelevant so-

cial cues can increase altruism. Many studies provide this kind of evidence, but some do not.

We conducted our study with children. Children are interesting subjects because they provide a means of addressing both the weak and strong versions of the evolutionary legacy hypothesis. With respect to the weak hypothesis, we did not find a pure treatment effect. With respect to the strong hypothesis, we found systematic variation in prosocial behaviour, with males being considerably less altruistic than females, but we did not find any associated variation in sensitivity to eye spots. As we discuss at length below, these experimental findings hold different implications for the weak and strong hypotheses. To summarize briefly, they refine but otherwise pose no real challenge to the weak hypothesis. A null result of any kind, including ours, simply shows that eye spots do not work under a specific set of conditions. For the strong hypothesis, however, limitations on eye spot effects suggest potential limitations on implicit reputation more broadly as a comprehensive explanation for anonymous one-shot altruism among unrelated strangers.

Overall, experimental results on eye spot effects are thoroughly mixed. Our null result constitutes one finding among a growing catalog of both positive and negative results. Two recent papers (Nettle *et al.*, 2013; Sparks and Barclay, 2013) presented exhaustive meta-analyses of eye spot studies to clarify the reasons for the mixed empirical results. Two main conclusions emerged. First, eye spots do not reliably affect the average degree of generosity, but they do seem to increase the probability a person gives some positive amount (Nettle *et al.*, 2013). Our data show no effect of either sort. Whether we analyze the magnitude of dictator transfers or simply the probability of giving, eye spots did not affect dictator choices in our experiment.

The second general conclusion to emerge from recent meta-analyses is that eye spot effects are conditional. Specifically, Sparks and Barclay (2013) concluded that eye spots tend to have a reliable effect only when exposure is sufficiently brief and shortly before the decision maker chooses. In our case, the experiment was not computerized. Consequently, for decisions to be anonymous, a face-like stimulus was continuously present for each subject inside the tent where the experiment took place. Exposure necessarily lasted the entire 5-10 minutes required for each

participant. The relevant exposure threshold, which is not known precisely, may be less than this (Sparks and Barclay, 2013). Thus, the children in our study may not have responded to eye spots because exposure time was longer than the critical threshold, whatever this threshold may finally prove to be.

The exposure time result suggests two general possibilities that are relevant for interpreting both our results and the results of other studies that expose participants to payoff-irrelevant reputational cues. First, cues may only have an effect if a number of conditions are met, and the requisite conditions could vary from one cue to another. Second, this kind of conditionality might limit the scope for implicit reputation to provide a stand-alone evolutionary explanation for anonymous one-shot altruism.

To illustrate, we did not find an eye spot effect. More precisely, we did not find an effect using the specific cue we used with an exposure time of 5-10 minutes among children of ages five and eight playing a dictator game in German-speaking schools in Switzerland. Change one or more of these details, and perhaps we would have had a different result. All we can say is the following. In an effort to focus on the consequences of working with children as subjects, we reproduced a number of methodological details that have been important for producing positive effects in the past (Haley and Fessler, 2005). These details include both our use of ear muffs and our use of the same stylized face in Haley and Fessler (2005). The fact remains, however, that a number of methodological details were probably unique to our study, and any one of them could have prevented an eye spot effect. For the most part, the conditions necessary for reputational cues to affect behavior have not been systematically studied, and collectively we know little about when such cues should matter.

Accordingly, a long-term research program that systematically unpacks reputational cues would be a fine undertaking. Indeed, our guess is that such a program will emerge, and it will lead to an extensive repository of new experimental methods and associated results. Whatever the details of a such a research program might turn out to be, ideally they will provide us with a detailed and thorough understanding of when implicit reputational concerns actively affect behavior. Importantly, because the weak hypothesis is silent about the importance of other evolutionary forces, the weak hypothesis would be fully consistent with

such a research program and its attempt to map the conditions controlling evoked reputational concerns. Indeed, from the perspective of the weak hypothesis, knowing when evoked reputation does and does not affect altruism is precisely what we would like to know; the question is intrinsically interesting.

The strong hypothesis requires something more extreme. In particular, if payoff-irrelevant reputational cues only increase altruism under specific conditions, then we need to know if all the relevant conditions are met in all observed instances of anonymous one-shot altruism among unrelated strangers. If so, then we have no reason to reject the strong hypothesis. If not, then we are left with some measure of altruism to explain. To show what we mean, we consider our own null result.

As discussed above, protracted exposure time could explain why the eye spots we used did not affect behavior. If a sufficiently brief exposure time is critical for an effect, however, in our experiment or any other experiment, then this requirement could limit the scope for evoked reputation and reciprocity to provide a complete explanation for anonymous one-shot altruism. For example, the economic experiments providing much of the evidence for anonymous one-shot altruism typically involve uncontrolled stimuli that could trigger an ancestral, reciprocity-based psychology. Subjects often mingle outside the lab before entering, they participate in a lab full with other people, they often interact with the experimenters, and they know they might see at least one or two of their fellow participants on the way out the door.

In many experiments, exposure to all these uncontrolled stimuli lasts for several minutes or even much longer, and participants could habituate to all reputational cues in this period of time. If they do habituate in this way, the cues cannot be producing observed altruism. Sparks and Barclay (2013) suggest that this is an important possibility because they show that habituation is often surprisingly quick. On the other hand, habituation times could vary greatly by cue. People may habituate rapidly to stylized faces, but they may only habituate slowly to the presence of other people in the room. Slow habituation for some cues would mean that implicit reputation could in fact be important in typical experimental settings. We cannot say, and indeed to our knowledge no one has good evidence on subtleties of this sort. We simply want to make a general point. Restrictions

on the activation of a reputational psychology could limit the scope, perhaps dramatically so, for reputation to provide a complete explanation for anonymous one-shot altruism. In the end, only the empirical details will tell.

The strong hypothesis potentially faces an analogous bind if we consider the degree of anonymity in an experiment. Much like Haley and Fessler (2005), we attempted to achieve considerable anonymity for our participants. We used both a tent and sound-proof ear muffs, for example, to eliminate uncontrolled social stimuli. The logic behind such an approach is to isolate and maximize the treatment difference. If social stimuli matter, and if both treatments include a lot of uncontrolled social stimuli, the treatment difference becomes relatively trivial and thus impossible to detect. Some researchers, however, have suggested that too much anonymity can diminish and even eliminate the response to eye spots and similar stimuli (Lamba and Mace, 2010; Raihani and Bshary, 2012). If so, this could explain why we did not observe a treatment effect. Like exposure time, however, this explanation also suggests limitations on the explanatory scope of the strong hypothesis. Eye spot effects have become an influential contribution in part because they apply conventional evolutionary mechanisms associated with repeated interactions to anonymous one-shot altruism (Burnham and Johnson, 2005), a behavior that is otherwise hard to understand. If too much anonymity destroys the effect, however, in our experiment or any other experiment, then eye spots are not working in precisely those situations where observed altruism is generating the greatest scientific controversy.

Apart from conditional eye spot effects, heterogeneity in altruism further illustrates the differences between the weak and strong hypotheses. We know, for example, that boys are more impulsive than girls (Chapple and Johnson, 2007), and indeed we observed that boys were less altruistic than girls in our experiment. The issue at hand is understanding what this kind of variation means more generally when interpreting results from an eye spot experiment. If variation of this sort is independent of an evoked reputation, it will increase variation within treatment groups. This can swamp any eye spot effect and make the effect hard to detect without proper statistical controls. This kind of situation does not generally undermine the weak hypothesis. If multiple factors independently affect anonymous one-shot altruism, they can reduce power, but they do not imply that

evoked reputational concerns never matter.

The strong hypothesis, however, does not admit forces that are independent of evoked reputation. All relevant factors must ultimately influence anonymous one-shot altruism via the activation of a reputational psychology. Thus, if the strong hypothesis is true, the only way to reduce power is to have many uncontrolled cues of observability that increase behavioral variation within treatments. We tightly controlled cues of observability specifically to avoid this kind of problem. The upshot is the following. Multiple factors and low power might be a plausible explanation for a null result in an eye spot experiment even if the weak hypothesis is true. This explanation, however, is less compelling under the strong hypothesis precisely because the strong hypothesis severely restricts the set of mechanisms that can generate uncontrolled variation in anonymous one-shot altruism.

Ultimately, we suspect that the challenges facing eye spot research will be interpretive. The empirical results are subject to refinement (Nettle *et al.*, 2013; Sparks and Barclay, 2013). Some studies will produce positive results; some will produce negative results. On-going research will sort out the differences. Whatever these differences turn out to be, the number of existing studies with positive results suggests that implicit reputational effects in some form are here to stay. Nonetheless, the accumulation of empirical studies may reveal that reputational cues only affect behavior under a restricted set of conditions. The key task will be both to identify the relevant conditions and to decide if they leave us with any residual altruism to explain. If not, researchers can proceed apace with the strong hypothesis and its appeal to theoretical convention. Otherwise, we might be left with our original challenge of how to explain altruism in situations that are anonymous and one-shot, both explicitly and implicitly.

Acknowledgements

We would like to acknowledge the generous support of the Swiss National Science Foundation (Grant No. 100014-130127/1, “The Social Dynamics of Normative Behavior”) and the European Research Council (Grant No. 295642, “Foundations of Economic Preferences”). We would also like to thank Jonas Blatter, Matthias Fehlmann, and Johanna Espin for their excellent research assistance.

References

- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, editors, *Second International Symposium on Information Theory*, pages 267–281. Budapest: Akademiai Kiado.
- Axelrod, R. and Hamilton, W. D. (1981). The evolution of cooperation. *Science*, **211**(27 March), 1390–1396.
- Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, **2**(3), 412–414.
- Bell, A. V., Richerson, P. J., and McElreath, R. (2009). Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proceedings of the National Academy of Sciences*, **106**(42), 17671–17674.
- Benenson, J. F., Pascoe, J., and Radmore, N. (2007). Children’s altruistic behavior in the dictator game. *Evolution and Human Behavior*, **28**(3), 168–175.
- Blake, P. R. and McAuliffe, K. (2011). i had so much it didnt seem fair: Eight-year-olds reject two forms of inequity. *Cognition*, **120**(2), 215 – 224.
- Bowles, S. (2004). *Microeconomics: Behavior, Institutions, and Evolution*. New York: Russell Sage.
- Bowles, S. (2006). Group competition, reproductive leveling, and the evolution of human altruism. *Science*, **314**(8 December), 1569–1572.
- Bowles, S. (2009). Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science*, **324**(5932), 1293–1298.
- Bowles, S. and Gintis, H. (2011). *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton: Princeton University Press.
- Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
- Boyd, R. and Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. Oxford: Oxford University Press.

- Boyd, R., Richerson, P. J., and Henrich, J. (2011). Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral Ecology and Sociobiology*, **65**(3), 431–444.
- Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag, 2nd edition.
- Burnham, T. C. (2013). Toward a neo-darwinian synthesis of neoclassical and behavioral economics. *Journal of Economic Behavior & Organization*, **90**, **Supplement**(0), S113 – S127. Evolution as a General Theoretical Framework for Economics and Public Policy.
- Burnham, T. C. and Hare, B. (2007). Engineering Human Cooperation. *Human Nature*, **18**(2), 88–108.
- Burnham, T. C. and Johnson, D. D. P. (2005). The Biological and Evolutionary Logic of Human Cooperation. *Analyse & Kritik*, **27**(2), 113–135.
- Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.
- Carbon, C.-C. and Hesslinger, V. M. (2011). Bateson et al.’s (2006) Cues-of-Being-Watched Paradigm Revisited. *Swiss Journal of Psychology*, **40**(4), 203–210.
- Chapple, C. L. and Johnson, K. A. (2007). Gender Differences in Impulsivity. *Youth Violence and Juvenile Justice*, **5**(3), 221–234.
- Cosmides, L. and Tooby, J. (2013). Evolutionary Psychology: New Perspectives on Cognition and Motivation. *Annual Review of Psychology*, **64**, 201–229.
- Ekström, M. (2012). Do watching eyes affect charitable giving? evidence from a field experiment. *Experimental Economics*, **15**(3), 530–546.
- Engelmann, J. M., Herrmann, E., and Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS ONE*, **7**(10), e48433.

- Engelmann, J. M., Over, H., Herrmann, E., and Tomasello, M. (2013). Young children care more about their reputation with ingroup members and potential reciprocators. *Developmental Science*, **16**(6), 952–958.
- Ernest-Jones, M., Nettle, D., and Bateson, M. (2011). Effects of eye images on everyday cooperative behavior: a field experiment. *Evolution and Human Behavior*, **32**(3), 172 – 178.
- Fehr, E. and Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, **8**(4), 185–190.
- Fehr, E. and Henrich, J. (2003). Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In P. Hammerstein, editor, *Genetic and Cultural Evolution of Cooperation*, pages 55–82. Cambridge: The MIT Press.
- Fehr, E. and Schneider, F. (2010). Eyes are on us, but nobody cares: are eye cues relevant for strong reciprocity? *Proceedings of the Royal Society B: Biological Sciences*, **277**(1686), 1315–1323.
- Fehr, E., Bernhard, H., and Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, **454**(28 August), 1079–1084.
- Francey, D. and Bergmüller, R. (2012). Images of Eyes Enhance Investments in a Real-Life Public Good. *PLoS ONE*, **7**(5), e37397.
- Hagen, E. H. and Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology*, **69**(3), 339 – 348.
- Haley, K. J. and Fessler, D. M. T. (2005). Nobody’s watching?: subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, **26**(3), 245–256.
- Hamilton, W. D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology*, **7**(1), 1–52.
- Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization*, **53**(1), 3–35.

- Henrich, J., Robert, B., Bowles, S., Camerer, C., Fehr, E., and Gintis, H. (2004). *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., and Ziker, J. (2006). Costly punishment across human societies. *Science*, **312**(5781), 1767–1770.
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., and Ziker, J. (2010a). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, **327**(5972), 1480–1484.
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010b). The weirdest people in the world? *Behavioral and Brain Sciences*, **33**, 61–83.
- House, B. R., Silk, J. B., Henrich, J., Barrett, H. C., Scelza, B. A., Boyette, A. H., Hewlett, B. S., McElreath, R., and Laurence, S. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences*, **110**(36), 14586–14591.
- Kagel, J. H. and Roth, A. E. (1995). *The Handbook of Experimental Economics*. Princeton: Princeton University Press.
- Kurzban, R. (2010). *Why everyone (else) is a hypocrite*. Princeton: Princeton University Press.
- Lamba, S. and Mace, R. (2010). People recognise when they are really anonymous in an economic game. *Evolution and Human Behavior*, **31**(4), 271 – 278.
- Lehmann, L., Feldman, M. W., and Foster, K. R. (2008). Cultural transmission can inhibit the evolution of altruistic helping. *The American Naturalist*, **172**(1), 12–24.
- Mifune, N., Hashimoto, H., and Yamagishi, T. (2010). Altruism toward in-group members as a reputation mechanism. *Evolution and Human Behavior*, **31**(2), 109 – 117.

- Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., and Bateson, M. (2013). The watching eyes effect in the Dictator Game: it's not how much you give, it's being seen to give something. *Evolution and Human Behavior*, **34**(1), 35–40.
- Nowak, M. A. and Sigmund, K. (1998). The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, **194**(4), 561–574.
- Oda, R., Niwa, Y., Honma, A., and Hiraishi, K. (2011). An eye-like painting enhances the expectation of a good reputation. *Evolution and Human Behavior*, **32**(3), 166 – 171.
- Panchanathan, K. and Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, **432**(25 November 2004), 499–502.
- Raihani, N. J. and Bshary, R. (2012). A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game. *Proceedings of the Royal Society B*, **279**(1742), 3556–3564.
- Richerson, P. J. and Boyd, R. (2005). *Not By Genes Alone: How Culture Transformed the Evolutionary Process*. Chicago: University of Chicago Press.
- Rigdon, M., Ishii, K., Watabe, M., and Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology*, **30**(3), 358–367.
- Smith, E. A. (2000). Three styles in the evolutionary analysis of human behavior. In L. Cronk, N. Chagnon, and W. Irons, editors, *Adaptation and Human Behavior: An Anthropological Perspective*, pages 27–46. New York: Aldine de Gruyter.
- Sparks, A. and Barclay, P. (2013). Eye images increase generosity, but not for long: the limited effect of a false cue. *Evolution and Human Behavior*, **34**(5), 317–322.
- Tane, K. and Takezawa, M. (2011). Perception of Human Face Does Not Induce Cooperation in Darkness. *Letters on Evolutionary Behavioral Science*, **2**(2), 24–27.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, **46**(1), 35–57.

Verbeek, M. (2008). *A Guide to Modern Econometrics*. West Sussex, England: John Wiley & Sons, Ltd, 3rd edition.

Williams, G. C. (1966). *Adaptation and Natural Selection*. Princeton: Princeton University Press.

Table 1: Ordinal logistic model of transfers. Independent variables indicate if the dictator dropped the recipient's envelope in the sealed box and also if she took her own envelope when exiting the tent. The independent variables additionally indicate the age (Primary), sex (Male), and treatment (Eye Spot) for each dictator. The best fitting model (electronic supplementary material) does not include the treatment dummy. The number of observations is 177.

Parameter	Odds ratio	Robust std. error	z	p
Not Dropped	2.344	1.159	1.72	0.085
Not Taken	1.614	0.903	0.86	0.392
Primary	0.598	0.194	-1.59	0.112
Male	0.421	0.119	-3.06	0.002
Eye Spot	0.791	0.226	-0.82	0.412

Intercepts	Estimate	Robust std. error
Intercept 0/1	-1.227	0.406
Intercept 1/2	-0.662	0.392
Intercept 2/3	-0.314	0.391
Intercept 3/4	0.222	0.392
Intercept 4/5	0.846	0.401
Intercept 5/6	1.544	0.412
Intercept 6/7	1.872	0.429
Intercept 7/8	2.112	0.435
Intercept 8/9	2.300	0.442

Table 2: Logistic model of transfers. Independent variables indicate if the dictator dropped the recipient’s envelope in the sealed box and also if she took her own envelope when exiting the tent. The independent variables additionally indicate the age (Primary), sex (Male), and treatment (Eye Spot) for each dictator. The best fitting model (electronic supplementary material) does not include the treatment dummy. The total number of observations is 177.

Parameter	Odds ratio	Robust std. error	z	p
Intercept	3.086	1.269	2.74	0.006
Not Dropped	1.156	0.505	0.33	0.739
Not Taken	2.756	2.311	1.21	0.227
Primary	0.931	0.349	-0.19	0.849
Male	0.372	0.126	-2.91	0.004
Eye Spot	0.851	0.277	-0.50	0.619

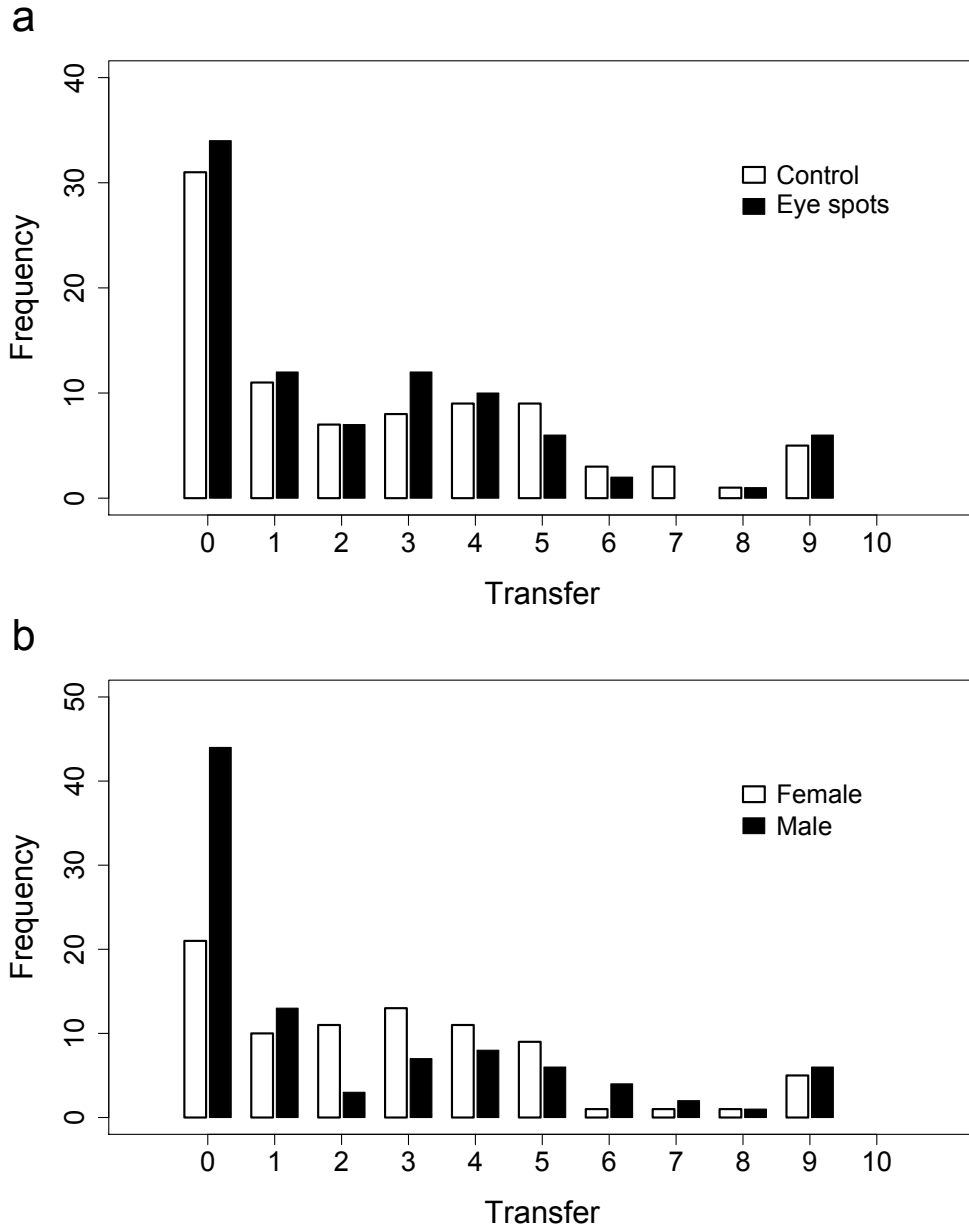


Figure 1: The distribution of dictator transfers. Dictator transfers are separated by treatment (a) and by the sex of the dictator (b). Model selection (electronic supplementary material) and associated regression results show that the sex of the dictator is an important variable for explaining variation in dictator transfers, while experimental treatment is not. In particular, boys transferred significantly less than girls (Table 1), and they transferred positive amounts at a significantly lower rate than girls (Table 2). The data consist of 177 observations.