

Original Communication

Bateson et al.'s (2006) Cues-of-Being-Watched Paradigm Revisited

Claus-Christian Carbon and Vera M. Hesslinger

University of Bamberg, Germany

Abstract. Bateson, Nettle, and Roberts (2006) provided an intriguing experimental paradigm for investigating the effects of social cues on cooperative behavior in a real-world setting. By placing an image of a pair of eyes on a cupboard door above an “honesty box” for hot beverages, they induced substantially higher amounts of contributions. As this finding has a significant impact on the social sciences and assumptions concerning the meaning of social cues for human behavior, we systematically reanalyzed their procedure and statistical analyses and tried to replicate the results while taking personality factors into account. The overall results of our analysis and replication efforts do not unequivocally support conclusions about the effects of eyes as social cues for cooperative behavior. Problems start with the definition of cooperative behavior, underspecified methods, confounding variables, invalid statistical analyses, and a lack of insight into the factors – including personality factors – that modulate the expected effect. A follow-up experiment with 138 participants showed no effect of eyes on socially relevant concepts and attitudes. Furthermore, none of the personality factors that, according to Bateson et al.'s explanation of reputational concerns, might be a source of effects interacted with any of the measures we used to operationalize these socially relevant concepts and attitudes.

Keywords: eyes, social cues, attitudes, face processing, cooperative behavior, personality traits, social desirability

A person's face contains a wealth of information. Even without further cues, we can assess another person's identity without great difficulty – except in some cases of so-called prosopagnosia (Grüter, Grüter, & Carbon, 2008). We can derive the emotional state and the current mood from specific activity patterns of facial muscles (Ekman & Friesen, 1975); detect sex, age, and race, attention status and responsiveness; and use facial expressions to reduce ambiguities in verbal signals (Bruce & Young, 1986). Thus, faces are of the utmost importance for social interaction. They can, partially or as a whole Gestalt, act as social cues (Haxby & Gobbini, 2007).

Bateson, Nettle, and Roberts (2006) presented an intriguing study model to demonstrate that perceiving even subtle facial cues (pictures of parts of faces, as in Leder & Carbon, 2005) might greatly change one's behavior. They attached the image of a pair of eyes to a cupboard door above an “honesty box” that was used to collect money to cover the costs of milk, coffee, and tea for the lab. To measure cooperative behavior, they ascertained the average amount of money participants put into the honesty box. Unfortunately, they did not explain *why* this behavior was considered cooperative when participants were paying for a good they had actually used, nor did they specify the extent to which these findings may generalize to other areas

of prosocial behavior. To test the selective impact of eyes as social cues on participants' behavior, they ran a control condition every other week in which they used equal-sized images of flowers instead of eyes as stimuli. After 10 weeks, the authors found an impressive contribution surplus of 176% for the eyes condition across all weeks. This extraordinarily clear result, which had a significant impact on the social cognition literature, calls for close inspection to assess the reliability of the effect.

The Present Research

In the present study, we aimed to systematically reanalyze Bateson et al.'s (2006) procedure and statistical analyses as well as the conclusions they drew. Furthermore, in an experimental study with 138 participants, we tested the impact of a pair of eyes on socially relevant concepts and attitudes. To test for personality factors potentially modulating the described effects, we also assessed participants on the Big Five personality dimensions (see Borkenau & Ostendorf, 1998). Bateson et al. monitored data collection only on a weekly basis and were, thus, unable to report how individual contribution behavior was affected by different

conditions. This was mainly a result of the use of a natural test setting, in which it is more difficult to implement optimal control conditions. In contrast, we randomly assigned our participants to the experimental or the control group.

Reanalysis of the Original Study by Bateson et al. (2006)

As a first step of critical reflection of the original study by Bateson et al. (2006), we reanalyzed central aspects of the paper, such as the procedure, stimulus material, sample characteristics, statistical analyses, and the inner logic of the paper including the operationalization of the dependent variable.

Analysis of the Procedure

Bateson et al. (2006) used a real-world setting to conduct their study. Unfortunately, their procedure for alternating between experimental conditions followed a strictly regular, nonrandomized order. They started with eyes for one week, followed by flowers for one week, followed by eyes again, and so on. The authors did not provide any information concerning contextual and situational specifics, so it is unclear whether the alternation between experimental conditions was confounded by factors such as periodic fluctuations in absenteeism or idiosyncratic paying behavior.

Furthermore, because the behavior of contributing to the honesty box was not linked to any individual person, we do not know who gave how much money how often. As a matter of fact, their procedure allowed them only to measure the total amount of money contributed per week, rendering it impossible to detect outliers, or even the distribution of individual cash transfers.

It is problematic that the authors related only the amount of money in the honesty box, which was intended to meet the expense of milk, tea, and coffee, to the amount of milk consumed. It is, thus, not clear whether the ratio of milk to tea and coffee consumption was, in fact, constant.

Analysis of the Stimulus Material

The authors developed a strong theoretical framework to explain their impressive effects by arguing that eye cues induce a feeling of being watched, which activates reputational concerns (Bateson et al., 2006, p. 413). These concerns are said to boost cooperative behavior. To reanalyze the original stimulus material, we conducted a rating study with 13 participants (7 males; $M_{\text{age}} = 33.2$ years, $SD = 10.5$). None of them participated in the main study, most were not linked to the university, and all were naïve to the purpose of the study and were either undergraduates or em-

ployed, those who were employed having diverse professions. We asked participants to rate the pictures in a sequential presentation task with unlimited time and employed the sequence used in the Bateson et al. (2006) study for the following variables: (1) "To what degree do the pictures appear to observe you?" and (2) "How frightening are the pictures?." We used these variables because these specific properties of the pictures used by Bateson et al. seem to be the most important triggers for prosocial behavior. All questions were answered on a 7-point Likert scale (1 = *not at all* to 7 = *very much*) without time pressure or a time limit. It took less than 5 min to complete the entire task. This study was not part of the main study. Both variables showed very high internal consistencies as measured by Cronbach's α , $\alpha_{\text{observing}} = .94$ and $\alpha_{\text{frightening}} = .98$. Although the ratings of the variable "To what degree do the pictures appear to observe you?" were, indeed, closely related to the amount of money put into the honesty box, Pearson $R = .81$, $p = .005$; the frightening quality was equally highly correlated to the amount of money contributed, $R = .80$, $p = .005$. So, on a merely statistical basis, we cannot conclude what the determinants of the higher contributions in the Bateson et al. study were.

Analysis of the Sample

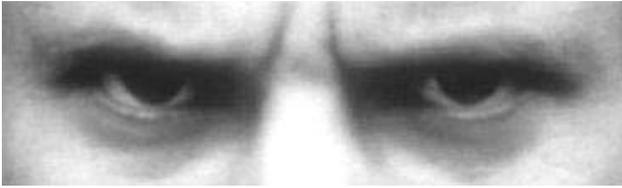
Bateson et al. (2006) described their sample very inaccurately, as they only reported the population size ($N = 48$), not the sample size. Consequently, it is not possible to estimate each individual's cash contribution, to investigate participants' behavior while they were processing the experimental displays, or to test the hypothesis on the degree to which the pictures appear to observe the participant while integrating theories concerning, for instance, typical personality moderators or gender effects. Again, we want to stress that this drawback is commonly encountered in field studies because of the demands of a natural context.

Analysis of the Statistical Analyses

The data are based on a single dependent measure assessed 10 times in a row. To analyze these data, the authors used a general linear model (GLM), which is not valid under the given conditions; for instance, as Bartlett's test showed an inequality of error variances, $\chi^2 = 121.8$, $p < .0001$. We graphically reanalyzed the data derived from Bateson et al.'s (2006) Figure 1 with the full range of statistical tests suitable for these data by comparing the distribution of contributions in the two experimental conditions (eyes vs. flowers). Of the tests employed, namely, the Mann-Whitney U, the Moses ($p = .103$), the Kolmogorov-Smirnov ($Z = 1.265$, $p = .082$), and the Wald-Wolfowitz test ($Z = -1.006$, $p = .167$), only the Mann-Whitney U test ($U = 1.0$, $Z = -2.402$, $p = .016$) revealed a significant effect.

Additionally, to obtain important information regarding

a) eyes



b) flowers



Figure 1. Stimulus material used in our experiment: (a) eyes for the social cue condition and (b) flowers for the neutral condition.

the distribution of the data and the quality of the GLM, the original GLM was conducted again. If we had ignored the premises needed for conducting an uncorrected GLM analysis, we would, indeed, have replicated Bateson et al.'s (2006) findings by showing a significant effect of Social Cue (eyes vs. flowers), $F(1, 7) = 11.82, p = .011, \eta_p^2 = 0.628$ – for purposes of direct comparison, the Bateson et al. analysis yields $F(1, 7) = 11.55, p = .011$. However, when we corrected the statistics according to the violated premises (e.g., the inequality of variances), we obtained a nonsignificant result, $F(2, 7) = 4.56, p = .054, ns$.

To further analyze the effects reported by Bateson et al. (2006), we conducted an experiment with 138 volunteers replicating parts of the original study while additional variables were integrated, most importantly personality dimensions and a variety of dependent measures.

Experiment

Participants

Participants ($N = 138$; 22 men; $M_{\text{age}} = 22.2$ years, $SD = 3.6$) were undergraduates volunteering for course credit. Three participants from the original sample of 141 had to be excluded because of missing personality dimension data. Participants were randomly assigned to either the experimental ($n = 69$; 9 men; $M_{\text{age}} = 22.3$ years, $SD = 3.5$) or the control group ($n = 69$; 13 men; $M_{\text{age}} = 22.1$ years, $SD = 3.6$). All participants were tested individually. With the experimental design used and the given n 's ($n_1 = 69$; $n_2 = 69$), an α of .20 and a statistical power of $\beta = .95$ (Faul, Erdfelder,

Buchner, & Lang, 2009), we are able to detect effect sizes of $d \geq .424$, which corresponds to a small to medium effect following Cohen's (1988) definition of effect sizes. Given the statistical power of 95%, which means that we will only find no effect at all although one exists in 5% of the cases, it is highly improbable that one would not detect a very large effect ($d = 1.948$; see Analysis of Statistical Analyses in the section reanalyzing the original study) as revealed by Bateson et al. (2006).

Materials

Personality Scale

To assess the personality dimensions following the NEO framework of the Big Five Inventory (BFI) of personality traits (e.g., Costa & McCrae, 1992), we employed the short version of the BFI (BFI-K; see Rammstedt & John, 2005), which has 21 items that characterize the five personality dimensions: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness.

Dependent Measures

To measure socially relevant concepts and attitudes, we created four scenarios for which the participants were to rate how probable their behavior in such a situation would be on a 7-point response scale ranging from 1 = *not at all* to 7 = *absolutely sure*. All four scenarios are outlined in Table 1.

Table 1

List of socially relevant scenarios: scenario number, name of the corresponding variable, and paraphrased contents of the scenario (English translation of the German version)

No.	Variable	Content
1	Prosocial-Immaterial	Imagine that your fellow student has been absent for 3 weeks. Do you help her/him to get up to date by lending him your notes and documents?
2	Prosocial-Material	Imagine that an older resident in your neighborhood has lost her/his job. It is also common knowledge that she/he is broke. Do you support her/him financially?
3	Revenge	Imagine that you're walking around downtown alone at night and cross paths with a group of young thugs. The leader of the pack starts insulting you and soon the others join in, too, and you feel intimidated. Two days later, you are downtown with your friends, and you run into the leader again, and he is now alone. Would you feel a need for revenge?
4	Help	Imagine again that you're having a stroll downtown at night. You see a group of drunkards who are intimidating a teenager who appears to be drunk. Others also see what is happening, but don't care. Will you try to help the boy stand up to the drunkards?

Lastly, participants estimated how often they wash their hands in the course of the day (variable *hand washing*) and indicated their attitude toward the idea of having a free will (“What do you think: How free is your will?,” i.e., variable *free will*) on a 7-point Likert scale (1 = *totally unfree* to 7 = *totally free*) following Stroessner and Green (1990).

Layout of Questionnaire

The material was combined to fill both sides of a DIN A4 (21 cm × 29.7 cm) page with the BFI-K on the front side and the rest on the back side. On the back side, we reserved spaces for presenting the image of the pair of eyes or that of the flowers, reflecting the experimental and the control condition, respectively (in the following referred to as the variable *social cue*). These spaces were positioned on the header and the footer and were 9.0 cm × 2.3 cm. We aimed to use images, for both experimental conditions, that had a similar overall structure and complexity. This was realized by using an image of two flowers positioned side by side at the same distance as the two eyes (see Figure 1). Also, because of Bateson et al.’s (2006) idea that people behave more cooperatively when they are being observed, we took great care to present eyes that were expected to be highly effective as social cues in the eyes condition.

Procedure

First, participants completed the BFI-K, which was employed at the beginning of the experiment to prevent biases from the experimental condition. Then they were to turn the page over to answer the remaining items. All items were processed consecutively; people were not allowed to go back to former items. The whole procedure lasted approximately 10 min.

Results and Discussion

Using a multivariate approach, we analyzed the impact of the experimental condition on the dependent variables from the socially relevant scenarios: *prosocial-immaterial*, *prosocial-material*, *revenge*, and *help*. Then we tested the impact on hand washing and free will with two separate univariate analyses. See Figure 2 for the mean data. These analyses were further qualified by integrating personality variables as measured by the BFI-K.

Impact on Socially Relevant Scenarios

To analyze the impact of the eyes on socially relevant scenarios, we conducted a multivariate analysis of variance (MANOVA) with the four socially relevant scenario vari-

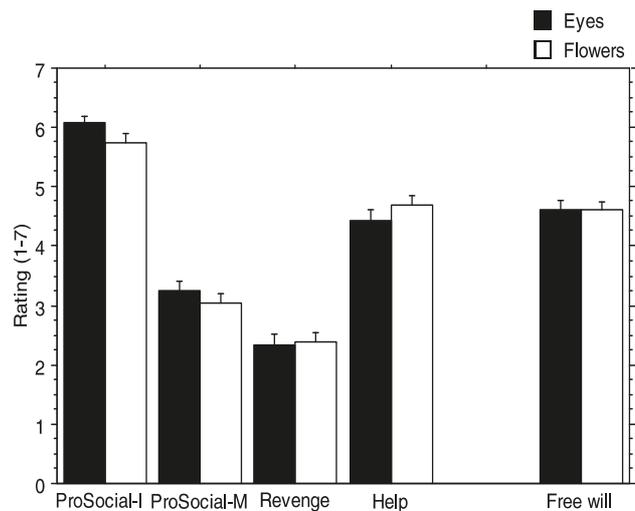


Figure 2. Impact of Social Cue on the socially relevant dependent variables Prosocial-Immaterial, Prosocial-Material, Revenge, and Help and the variable Free Will on the associated 7-point Likert scales. Error bars indicate 1 standard error of the mean.

ables (prosocial-immaterial, prosocial-material, revenge, and help) as dependent measures and social cue (eyes vs. flowers) as a between-participants independent measure. The variable social cue was not significant, $F(1, 133) < 1$, $p = .5618$, *ns*. Note that, as described in the subsection Participants, this experiment was configured to reduce the risk of falsely indicating null effects when, in fact, small to medium-sized effects, according to Cohen’s (1988) definition, are at work to 5% of all cases. Additionally, to be sure that no effect of social cue was neglected because of a noncoherent concept of attitudes toward socially relevant behavior, we decided to conduct four separate univariate ANOVAs for the scenario variables. None of the effects was found to be significant: $F_{\text{prosocial-immaterial}}(1, 136) < 1$, $p = .4859$, *ns*; $F_{\text{prosocial-material}}(1, 136) < 1$, $p = .5795$, *ns*; $F_{\text{revenge}}(1, 136) < 1$, $p = .4496$, *ns*; and $F_{\text{help}}(1, 136) = 1.95$, $p = .1653$, *ns*.

Impact on Hand Washing

To test the impact of social cue on an everyday behavior of prosocial relevance that is tightly linked with social desirability, we performed a univariate ANOVA with the variable hand washing ($M_{\text{eyes}} = 9.91$ vs. $M_{\text{flowers}} = 9.27$). Again, social cue had no significant influence, $F_{\text{hand washing}}(1, 136) < 1$, $p = .4010$, *ns*.

Impact on Free Will

We tested the impact of social cue on participants’ attitudes toward the idea of having a free will. This was done to

Table 2
 Statistics of multiple regression analyses with personality dimensions as predictors and the dependent variables as target variables; for better visibility, only the significant predictors are listed

Dependent variable	Predictor	Beta	F	p
Prosocial-Immaterial	Extraversion	.176	4.33	.0392
Prosocial-Material	Agreeableness	.198	5.53	.0201
Help	Agreeableness	.263	10.07	.0019
Hand-Washing	Conscientiousness	.203	5.80	.0174
Free Will	Neuroticism	-.247	8.80	.0036

assess a variable of utmost importance for humans as well as specific expectations for prosociality as evaluated by the previous variables. We performed a univariate ANOVA with the free will variable, but did not find an effect of social cue, $F_{\text{free will}}(1, 136) < 1, p = 1.0, ns$.

Testing for Personality Factors

Bateson et al. (2006) proposed that people show stronger cooperative behavior because of reputational concerns (p. 413). Accordingly, we should find participants with specific patterns of personality traits who are particularly susceptible to such effects (see Fehr & Schneider, 2010). For instance, people with a high level of Agreeableness should show increased prosocial behavior in the presence of a social cue such as a pair of eyes. Also, neurotic individuals are believed to be particularly responsive to social cues such as observing eyes. To test such influences, we conducted multiple regression analyses with the five personality dimensions as predictors and our dependent variables as target variables. As shown in Table 2, personality dimensions did have significant predictive power with regard to several of our dependent variables, documenting the importance of integrating personality dimensions into such a study. Furthermore, the specific and highly plausible pattern of predictors demonstrated the adequacy of our dependent measures. For instance, the higher Agreeableness was, the more the concept of helping behavior was activated. Moreover, participants with higher levels of Neuroticism were less likely to regard their will as being free (see Table 2).

Having shown that personality dimensions are valuable predictors of a variety of socially relevant variables, as well as of variables such as estimated frequency of hand washing (which is also linked to prosocial behavior since it reduces the risk of contagious diseases) and attitude toward one's perceived freedom of will, we tested whether these personality dimensions, in interaction with social cue, affected the dependent variables. This was done by a multivariate analysis of covariance for the four socially relevant variables and two univariate analyses of covariance for hand washing and free will. The sequence of these analyses

follows the logic of the (M)ANOVAs above, but, in line with the covariance approach, personality dimensions are also included. To test the hypothesis that social cue interacts with personality dimensions, we focused on two-way interactions between social cue and the covariates. None of the interactions for the socially relevant variables or for hand washing were significant, $F_{\text{socially relevant}}(8, 248) < 1.73, ps > .0924, ns$; $F_{\text{hand washing}}(2, 126) < 2.51, ps > .0853, ns$. For free will, we only found one significant interaction, namely between Neuroticism and social cue, $F(2, 126) = 4.67, p = .0110, \eta_p^2 = 0.069$. None of the other interactions were significant, $F_s(2, 126) < 1.36, ps > .2610, ns$. Thus, in contrast to Bateson et al.'s strong assumption that even weak social cues, such as a pair of eyes, trigger strong behavioral changes, we did not find any direct effect on any of our target variables. The only, very weak, interactive effect found involved a combination of the personality dimension neuroticism and one of the dependent variables, free will.

General Discussion

The present paper reflected on an important contribution to the social cognition literature made by Bateson et al. (2006). As a number of sources refer to Bateson et al.'s fundamental effect of even minimal social cues, such as eyes, causing strong changes in human behavior without criticizing the paper's general method, we replicated this condition using various approaches. First, we systematically reanalyzed their procedure and statistical analyses and the general operationalization they had in mind. We then designed a replication experiment with 138 participants who were also assessed in terms of personality factors to test the authors' hypotheses that social cues increase cooperative behavior.

The reanalysis revealed a series of shortcomings. Problems start with the definition of cooperative behavior, underspecified methods, confounding variables, invalid statistical analyses, and a lack of insight into the factors, including personality factors, modulating the described effect. The authors inferred that eyes as social cues are very effective in reprogramming human behavior, even under conditions different from those tested: "If even very weak, subconscious cues, such as the photocopied eyes used in this experiment, can strongly enhance cooperation, it is quite possible that the cooperativeness observed in other studies results from the presence in the experimental environment of subtle cues evoking the psychology of being observed" (p. 413).

With our replication experiment, we were not able to replicate the effect of eye cues affecting socially relevant concepts and attitudes, which is in accordance with a recently conducted study by Ekström (in press), who posted a picture of human eyes on recycling machines, at which the change dispensed could be donated. However, with a

Table 3
Selection of studies investigating the effects of subtle visual cues on participants' behavior. The "Findings" column provides insight into the replicability of the biasing effect of eyes on (social) behavior

Authors	Setting	Findings
Mifune, Hashimoto, & Yamagishi (2010)	Dictator game	Induction of in-group favoritism by display of painting of eyes on screen during game
Burnham & Hare (2007)	Public goods game	Enhancement of contributions to public good due to presentation of images of Kismet, a robot with human-like eyes, on screen during game
Rigdon, Ishii, Watabe, & Kitayama (2009)	Dictator game	Increase in giving behavior of male participants when three dots resembling a "watching-eyes" configuration are shown on screen during game; no corresponding increase for three dots in neutral configuration
Fehr & Schneider (2010)	Trust game	No impact of eye cues on trust behavior
Lamba & Mace (2010)	Ultimatum game	No impact of eye cues on offering behavior – neither for participants acting as proposer, nor for participants acting as responder
Ernest-Jones, Nettle, & Bateson (in press)	Field experiment on littering behavior	Substantial impact of eye cues on littering behavior: halving of the odds of littering in the presence of posters featuring eyes
Ekström (in press)	Field experiment on donation behavior	With a huge number of observations (16,775), no evidence for a social cue effect on donation behavior at recycling machines
Haley & Fessler (2005)	Economic game	Substantial increase in money given to game partners

very large number of observations (16,775; see Table 3), Ekström did not find any effect of displaying eyes versus displaying flowers. Furthermore, in our present study, we did not find a strong link between personality factors and the effect of eye cues on socially relevant concepts and attitudes. Following Bateson et al.'s (2006) explanation of reputational concerns as a source of effects, this might be relevant for the measures we used to operationalize these socially relevant concepts and attitudes (see Fehr & Schneider, 2010). It is also evident that the nature of the observing or frightening eyes does not seem to be the source of the effects documented by Bateson et al., as we used a cue particularly high on both dimensions. We could argue that Bateson et al. investigated real cooperative behavior as they measured a behavioral outcome – the amount of money in the honesty box – whereas we "only" looked at attitudinal data. Indeed, attitudes are not necessarily linked to subsequent behavior and sometimes attitude measures predict a totally different, even opposite, behavior ("effect of attitude-behavior inconsistency;" Zanna, Olson, & Fazio, 1980). However, this argument does not diminish the rele-

vance of our experiment. Webb and Sheeran (2006) conducted a rare meta-analysis in which they focused exclusively on experiments that tested the causal effects of attitudes on behavior. On the basis of 47 experimental studies, they concluded that medium-to-large changes in intention lead to small-to-medium changes in behavior. In the context of prosocial behavior, as focused on in Bateson et al.'s research, this would mean that many people have strong prosocial attitudes but show behavioral patterns that are, by comparison, less prosocial. This can be seen in everyday life with many phenomena linked to Bateson et al.'s honesty-box example, for instance, donation behavior (Anker, Feeley, & Kim, 2010): While people tend to announce large donations when first confronted with information about major catastrophes in the world reported by the media, final donation amounts are rarely high and sometimes nonexistent. Such findings should be observed on the basis of items or attitude changes that take place when people find themselves in a demanding situation entailing social desirability. For instance, hand washing behavior is an item that should be sensitive to biases resulting from social desirability. Typical prosocial behavior, also asked for in our experiment, should also be linked to social desirability. None of these, however, showed changes as a result of the presence of a pair of eyes. Additionally, we did not give any material goods for specific participants' responses. That no effect was found is a strong sign that the originally reported effect of social cues seems to be very small or highly fragile, as participants could have easily formed a good reputation for themselves without having to spend anything.

Furthermore, people with certain personality traits, for instance, people with a high level of Agreeableness, should show increased prosocial behavior in the presence of a social cue such as a pair of eyes (see Fehr & Schneider, 2010). It also seems rational to think of neurotic persons, who are more susceptible to social influences, as being particularly responsive to social cues such as observing eyes. Although we found clear links between personality factors and our dependent variables, which demonstrates the relevance of integrating such variables into an experimental design testing predictors for prosocial behavior, we hardly found any clear relationship between personality traits, experimental condition, and the dependent variables. Again, we interpret this result as contradicting Bateson et al.'s (2006) findings.

When qualifying our results via comparison with other studies that aimed to replicate the original effect, we observed inconsistent findings: Many of these studies did not find any effect of exposure to pairs of eyes on participants' behavior (see Table 3). Thus, our findings are in accordance with other studies that failed to replicate the effects of the original study.

Last, but not least, it is quite impressive how the system of the honesty box described by Bateson et al. (2006) worked for such a long time ("This system of payment for drinks had been in place for several years prior to the commencement of the current study," p. 412). If we calculate the money paid for a typical beverage consumed in the de-

scribed lab and compare it to the costs of these beverages, we are faced with an astonishing fact. Let us assume, as the authors did, that weeks with flowers reflected typical contribution behavior, namely, 0.15 GBP per liter of milk. Given that milk costs 0.53 GBP per liter on average, there is a negative balance of 0.38 GBP per liter. Based on the statistics provided by the International Coffee Organization (ICO) in 2008 and information about the world tea market in 2006 published by the United Nations Food and Agriculture Organization (FAO), we can conservatively estimate a daily consumption of one cup of coffee and three cups of tea per capita per day, plus the milk needed to whiten the beverages, namely, at least 0.025 liters per cup. Based on an average sickness absence rate of 1.7% (statistical data on absence rates in GB from the year 2005, Barham & Begum, 2005), the remaining staff (47 out of 48 persons) would consume 23.5 liters in total per week (5 work days) yielding a year's consumption of about 1,000 liters of milk. In sum, in 5 years the system would produce a deficit of more than 2,000 GBP.

So what does the original study by Bateson et al. (2006) tell us in the end? Bateson et al. demonstrated that faces or even parts of faces can play a crucial role in social behavior, although other sources (e.g., Ekström, in press), as well as the present paper, did not find such an effect. We know from everyday life, as well as from social interactions, from communication with others, and also from face identification, that eyes are highly important (Carbon & Leder, 2005). We also know from recent research that some of the frequently encountered statements that parts of faces play an essential role as detectors (Grüter & Carbon, 2010) and identifiers (Leder & Carbon, 2005) are not as empirically clear as formerly assumed. Bateson et al.'s findings seem to be very difficult to replicate. Furthermore, the study fails to mention important experimental details and control over some essential variables such as the composition of the sample, which makes any kind of replication very difficult from a technical perspective. Taken as a whole, the general idea that eyes *can play* an important role as social cues is indeed very important, but Bateson et al.'s (2006) publication does not seem to give the clearest evidence in favor of this idea.

Acknowledgments

We would like to thank the Women's Representatives of the University of Bamberg for a grant to VMH and the DAAD (Deutscher Akademischer Auslandsdienst) for a grant to CCC. We also want to express our gratitude to Sabine Albrecht for conducting the study, to Julia S. Fuchs and Michael "Wuschel" Groh for providing valuable sources for the manuscript, and to Ramona Luedtke and Andrea Lyman for proofreading the text. Last but not least, we would like to thank two excellent reviewers for their helpful and practical comments on an earlier version of the manuscript.

References

- Anker, A. E., Feeley, T. H., & Kim, H. (2010). Examining the attitude-behavior relationship in prosocial donation domains. *Journal of Applied Social Psychology, 40*, 1293–1324.
- Barham, C., & Begum, N. (2005). Sickness absence from work in the UK. *Office for national statistics. Labor market trends*. Retrieved from <http://www.statistics.gov.uk/hub/index.html>
- Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*, 412–414.
- Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Research in Personality, 32*, 202–221.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305–327.
- Burnham, T. C., & Hare, B. (2007). Engineering human cooperation: Does involuntary neural activation increase public goods contributions? *Human Nature: An Interdisciplinary Biosocial Perspective, 18*, 88–108.
- Carbon, C. C., & Leder, H. (2005). When feature information comes first! Early processing of inverted faces. *Perception, 34*, 1117–1134.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Costa, P. T., & McCrae, R. R. (1992). *NEO-PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.
- Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. Cambridge, MA: Malor.
- Ekström, M. (in press). Do watching eyes affect charitable giving: Evidence from a field experiment. *Experimental Economics*.
- Ernest-Jones, M., Nettle, D., & Bateson, M. (in press). Effects of eye images on everyday cooperative behavior: A field experiment. *Evolution and Human Behavior*.
- FAO, Food and Agriculture Organization of the United Nations. (2006). *FAO statistics division: Food supply*. Retrieved from <http://faostat.fao.org/site/609/DesktopDefault.aspx?PageID=609#ancor>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.
- Fehr, E., & 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*, 1315–1323.
- Grüter, T., & Carbon, C. C. (2010). Escaping attention: Some cognitive disorders can be overlooked. *Science, 328*, 435–436.
- Grüter, T., Grüter, M., & Carbon, C. C. (2008). Neural and genetic foundations of face recognition and prosopagnosia. *Journal of Neuropsychology, 2*, 79–97.
- Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior, 26*, 245–256.
- Haxby, J. V., & Gobbini, M. I. (2007). The perception of emotion and social cues in faces. *Neuropsychologia, 45*(1), 1.
- ICO, International Coffee Organization. (2008). *Historical coffee statistics*. London: Author. Retrieved from http://www.ico.org/new_historical.asp
- Lamba, S., & Mace, R. (2010). People recognize when they are

- really anonymous in an economic game. *Evolution and Human Behavior*, 31, 271–278.
- Leder, H., & Carbon, C. C. (2005). When context hinders! Learn-test compatibility in face recognition. *Quarterly Journal of Experimental Psychology. Section A*, 58, 235–250.
- Mifune, N., Hashimoto, H., & Yamagishi, T. (2010). Altruism toward in-group members as a reputation mechanism. *Evolution and Human Behavior*, 31, 109–117.
- Rammstedt, B., & John, O. P. (2005). Short version of the Big Five Inventory (BFI-K): Development and validation of an economic inventory for assessment of the five factors of personality. *Diagnostica*, 51, 195–206.
- Rigdon, M., Ishii, K., Watabe, M., & Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology*, 30, 358–367.
- Stroessner, S. J., & Green, C. W. (1990). Effects of belief in free will or determinism on attitudes toward punishment and locus of control. *Journal of Social Psychology*, 130, 789–799.
- Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, 132, 249–268.
- Zanna, M. P., Olson, J. M., & Fazio, R. H. (1980). Attitude-behavior consistency: An individual difference perspective. *Journal of Personality and Social Psychology*, 38, 432–440.

Claus-Christian Carbon

University of Bamberg
Department of General Psychology and Methodology
Markusplatz 3
DE - 96047 Bamberg
Germany
ccc@experimental-psychology.com