Evolution and Human Behavior 30 (2009) 212–224

Evolution and Human Behavior

No enhanced recognition memory, but better source memory for faces of cheaters☆

Axel Buchner*, Raoul Bell, Bettina Mehl, Jochen Musch

Department of Experimental Psychology, Heinrich-Heine-Universität Düsseldorf, D-40225 Düsseldorf, Germany

Initial receipt 31 March 2008; final revision received 26 January 2009

Abstract

Previous studies sought to test for the existence of a "cheater-detection module" by testing for enhanced memory for the faces of cheaters, but past results have been inconclusive. Here, we present four experiments showing that old–new discrimination was not affected by whether a face was associated with a history of cheating, trustworthy or irrelevant behavior. In contrast, source memory for faces associated with a history of cheating (i.e., memory for the cheating context in which the face was encountered) was consistently better than source memory for other types of faces. This pattern held under a variety of conditions, including different types of judgments participants made about the stimulus persons (attractiveness in Experiment 1; likeability in Experiments 2–4), different retention intervals (a few minutes in Experiments 1, 2 and 4; 1 week in Experiment 3), whether the behaviors were exceptional or ordinary (Experiments 1–3) and whether the social status of the characters was low or high (Experiment 4). Given no differences in old–new discrimination, enhanced source memory for faces of cheaters may be useful for avoiding cheaters in future interactions.

Keywords: Cheater detection; Human memory; Source memory

1. Introduction

Social cooperation is a universal feature of human societies and groups that may have evolved because individuals can increase their fitness by cooperating with each other (Axelrod & Hamilton, 1981; Cosmides & Tooby, 1989; Trivers, 1971). However, cooperation is also risky.

Some individuals may exploit their social-exchange partners by benefiting from them, but failing to reciprocate. Therefore, a strategy in which individuals cooperate regardless of the behavior of their exchange partners cannot be successful in the long run and would be replaced by more egoistic strategies. Cooperative strategies are only evolutionarily stable if they are accompanied by cognitive mechanisms that enable the individual to detect and avoid cheaters in social interactions (Axelrod & Hamilton, 1981; Trivers, 1971). Based on these considerations, it has been suggested that specialized modules have evolved within the human mind that help us to deal with social-exchange situations. Specifically, social contract theory (Cosmides, 1989; Cosmides & Tooby, 1989; Tooby & Cosmides, 2005) postulates brain mechanisms to have been selected during human evolution that are functionally specialized in the detection of cheaters. Integrated into a "cheater-detection module," these mechanisms supposedly allow the individual to quickly and easily draw inferences on whether someone has cheated in prior exchanges or is about to cheat in future interactions. Mealey, Daood, and Krage (1996) derived from this theory the prediction that faces of cheaters should also be remembered better than faces associated with other types of behavior. Indeed, it seems evident that avoiding potential cheaters based on memory for their previous behaviors may be of considerable benefit because harm can be avoided before it occurs. Mealy et al. reported that old–new discrimination of faces varied as a function of whether the depicted persons were described as cheaters, as trustworthy or in a way that was irrelevant to the cheating–trustworthiness dimension. For faces associated with low-status

☆ The research reported in this article was supported by a grant from the Deutsche Forschungsgemeinschaft (Bu 945/6–1).

* Corresponding author. Institut für Experimentelle Psychologie, Heinrich-Heine-Universität, D-40225 Düsseldorf, Germany.

E-mail address: axel.buchner@uni-duesseldorf.de (A. Buchner).

1090-5138/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.evolhumbehav.2009.01.004

professions, old–new discrimination was better for cheaters than for faces of people described as trustworthy. Unexpectedly, the pattern was reverse for faces associated with high-status professions. In both cases, old–new discrimination was intermediate for faces with descriptions that were irrelevant to the cheating–trustworthiness dimension. The authors interpreted their results as evidence of better memory for faces of cheaters. This finding is referred to in many evolutionary psychology handbooks and textbooks (e.g., Burnstein, 2005; Buss, 2004; Cartwright, 2000; Cummins, 2005; Gaulin & McBurney, 2001; Palmer & Palmer, 2002) and has been interpreted as evidence in favor of social contract theory (e.g., Buss, 2004; Cartwright, 2000; Mealey et al., 1996).

However, more recent studies using carefully controlled materials have been unable to replicate the Mealey et al. (1996) finding (Barclay & Lalumière, 2006; Mehl & Buchner, 2008). These findings are not problematic for social contract theory because it was inappropriate from the outset to focus selectively on old–new discrimination. This is so because improved old–new face discrimination per se, that is, just perceiving a face as familiar, cannot help avoiding cheaters and thus cannot provide an evolutionary benefit as long as the source or context in which the face had been encountered is not remembered concurrently. Even worse, greater familiarity of faces of cheaters without context information might increase the risk of being exploited because of the preference often exhibited towards familiar stimuli (Bornstein, 1989; Zajonc, 1968). In any case, given the finding of no difference in old–new discrimination (Barclay & Lalumière, 2006; Mehl & Buchner, 2008), social contract theory may allow deriving the prediction that source memory is improved for individuals with a history of cheating relative to individuals encountered in other situations. A source memory advantage for cheaters should be instrumental in avoiding cheaters and thus should be beneficial to socially cooperating individuals and groups.

Here we present four experiments designed to test the possibility of a source memory advantage for cheaters. The experiments followed the basic design of those reported by Mehl and Buchner (2008) which were modeled after the original experiment of Mealey et al. (1996). In an exposition phase, participants rated the attractiveness (Experiment 1) or likeability (Experiments 2, 3 and 4) of facial photographs presented together with descriptions of the depicted person's behavioral history. In a test phase, previously seen and new faces were judged as old or new. New in the present experiments, if a face was judged as old, participants indicated whether they thought that the person had been characterized by a history of cheating, of trustworthiness or by neither of these. We expected to replicate earlier findings that old–new discrimination does not differ for faces associated with different types of behavior. Our central hypothesis, however, was that if there was any validity in the derivation from social contract theory that humans have a specialized module for remembering cheaters, then source memory for faces characterized as cheaters should be better than source memory for faces associated with other behavior descriptions.

2. Measuring source memory

A problem when measuring memory for source is which measurement tool to use. Early approaches relied on ad hoc measures which confound old–new discrimination (item memory) with source memory and guessing processes (e.g., see the discussion of the conditional source identification measure, or CSIM, in Bayen, Murnane, & Erdfelder, 1996; Murnane & Bayen, 1996). Fortunately, alternative measurement tools exist in terms of multinomial models[1] of source memory (Batchelder, Hu, & Riefer, 1994; Batchelder & Riefer, 1990; Bayen et al., 1996; Hu & Batchelder, 1994; Riefer, Hu, & Batchelder, 1994). Compared to more conventional approaches, these models may appear slightly more complex at first sight, but they are nevertheless to be preferred because they have many advantages over other approaches to the analysis of source memory data. One important advantage is that multinomial models of source memory allow for the independent measurement of old–new discrimination, source memory and various types of guessing processes. We therefore analyzed the source memory data of the present experiments using the multinomial source memory model developed and successfully validated by Bayen et al. (1996). This model has been used successfully in a number of experiments (e.g., Bayen, Nakamura, Dupuis, & Yang, 2000; Bell, Buchner, & Mund, 2008; Dodson & Shimamura, 2000; Ehrenberg & Klauer, 2005; Simons et al., 2002; Spaniol & Bayen, 2002). An adaptation of the model for the present purposes is presented in Fig. 1.

The model displayed in Fig. 1 contains 12 parameters. Each parameter represents the probability with which certain cognitive processes occur. Parameter D_{Cheat} represents the probability of recognizing a cheater face shown in the exposition phase as old. Parameter d_{Cheat} represents the conditional probability of remembering correctly that a recognized face was encountered in the context of a history-of-cheating description. If the source of a correctly recognized face is not known (with probability $1-d_{\mathrm{Cheat}}$), then the correct history-of-cheating source may still be guessed with probability $a_{\mathrm{CheatTrust}} \cdot a_{\mathrm{Cheat}}$. Alternatively, it may be guessed (incorrectly) that the face is that of a person described as trustworthy with probability $a_{\mathrm{CheatTrust}} \cdot (1-a_{\mathrm{Cheat}})$. Finally, it may be guessed (again incorrectly) that the face is that of a person described as neither cheating nor trustworthy with probability $(1-a_{\mathrm{CheatTrust}})$. If a cheater face from the

[1] Historically, multinomial models seem to have their roots in statistical genetics where such models were used to infer gene frequencies from phenotypic category frequencies, such as the well-known multinomial model for the ABO blood group (Bernstein, 1925).

exposition phase is not correctly recognized as old (with probability $1{-}D_{\mathrm{Cheat}}$), it may still be guessed, with probability b, that the face is old. For these faces, the correct history-of-cheating source may be guessed with probability $g_{\mathrm{CheatTrust}}{\cdot}g_{\mathrm{Cheat}}$. Alternatively, it may be guessed (incorrectly) that the face is that of a trustworthy person with probability $g_{\mathrm{CheatTrust}}{\cdot}(1{-}g_{\mathrm{Cheat}})$. Finally, it may be guessed (incorrectly) that the face is that of a person described as neither cheating nor trustworthy with probability $(1{-}g_{\mathrm{CheatTrust}})$. The final branch in this tree of the model concerns cheater faces shown in the exposition phase that are neither recognized as old (with probability $1{-}D_{\mathrm{Cheat}}$) nor guessed to be old (with probability $1{-}b$). These faces are incorrectly judged to be new. Analogous considerations hold for the model trees for the other types of faces.

To illustrate the utility of the model consider, for instance, correct classifications of cheater faces as "cheaters." These classifications may be arrived at by recognizing the face as old and remembering its source (with probability $D_{\mathrm{Cheat}}{\cdot}d_{\mathrm{Cheat}}$), by recognizing the face as old and guessing its source [with probability $D_{\mathrm{Cheat}}{\cdot}(1{-}d_{\mathrm{Cheat}}){\cdot}a_{\mathrm{CheatTrust}}{\cdot}a_{\mathrm{Cheat}}$] or by guessing that the face is old and guessing its source [with probability $(1{-}D_{\mathrm{Cheat}}){\cdot}b{\cdot}g_{\mathrm{CheatTrust}}{\cdot}g_{\mathrm{Cheat}}$]. Thus, the probability of a cheater face receiving a "cheater" classification is given by $D_{\mathrm{Cheat}}{\cdot}d_{\mathrm{Cheat}}{+}D_{\mathrm{Cheat}}{\cdot}(1{-}d_{\mathrm{Cheat}}){\cdot}a_{\mathrm{CheatTrust}}{\cdot}a_{\mathrm{Cheat}}{+}(1{-}D_{\mathrm{Cheat}})\,b{\cdot}g_{\mathrm{CheatTrust}}{\cdot}g_{\mathrm{Cheat}}$. In other words, correct cheater

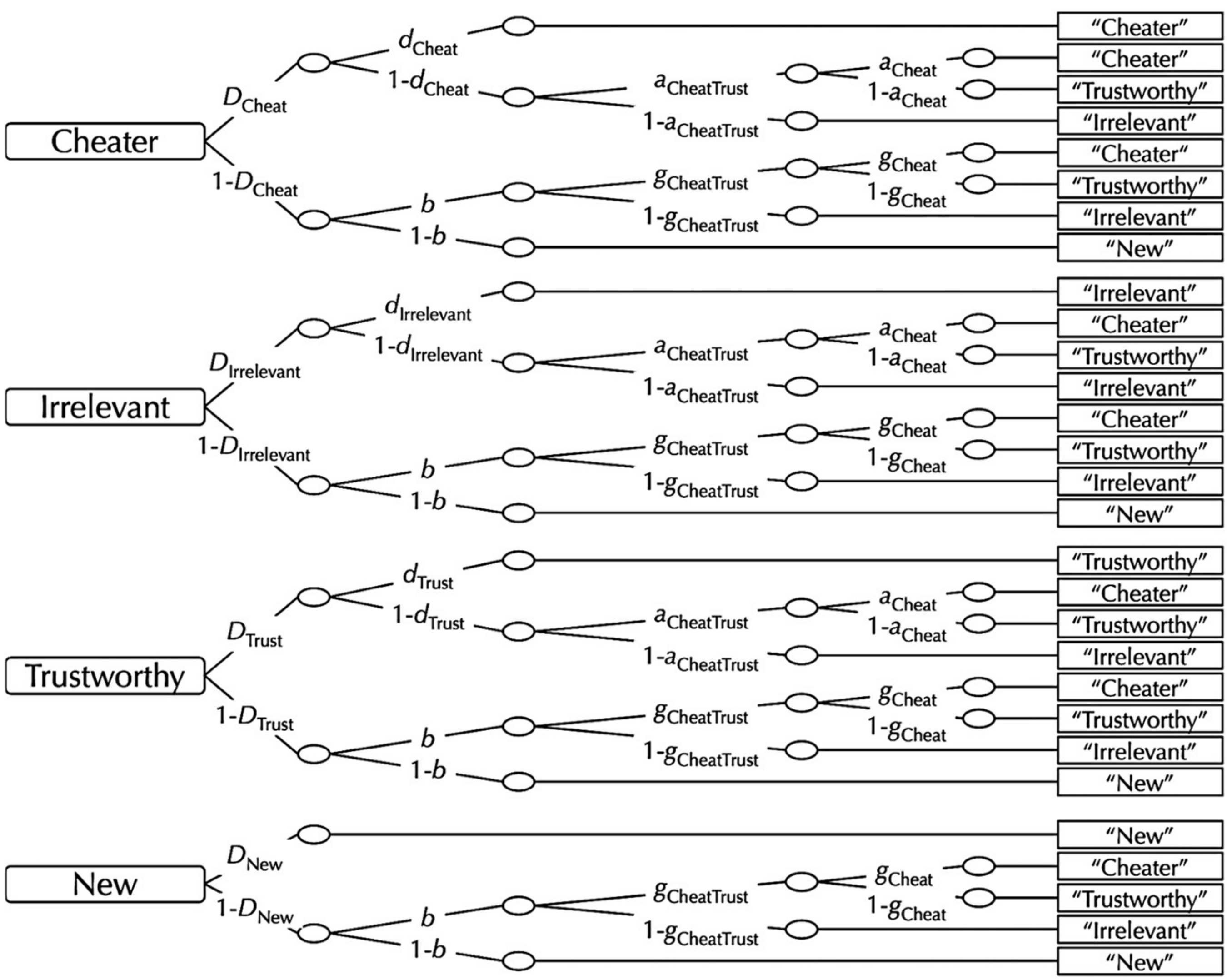

Fig. 1. Source memory model of Bayen et al. (1996) as adapted for the present purposes. Rectangles on the left side represent the types of faces presented (Cheater: face associated with cheating; Irrelevant: face associated with irrelevant behavior; Trustworthy: face associated with trustworthiness; New: new faces not presented in the exposition phase). Letters along the links represent the probabilities with which certain cognitive states occur [D: probability of correctly identifying a face as old (for previously presented faces) or new (for new faces); d: probability of correct source memory (i.e., remembering the context of encountering a face that was detected as old); b: probability of guessing that a nonrecognized face is old; a: probabilities of guessing that a recognized face for which the source was not remembered was encountered in a certain context; g: probabilities of guessing that a nonrecognized face that was guessed to be old was encountered in a certain context]. Rectangles on the right side represent the categories of participants' judgments.

classifications can be arrived at based on many different types of processes that remain hidden if one analyzes raw classification data. By using a multinomial model of source memory, it is possible to decompose classification performance into the processes involved and to estimate the probabilities associated with the model parameters representing these processes using standard computer programs (e.g., Rothkegel, 1999; Stahl & Klauer, 2007). What is more, statistical tests can be performed on these model parameters. For instance, the hypothesis that source memory is better for cheater faces than for trustworthy faces implies that the model parameter representing source memory for cheaters, d_{Cheat}, is larger than the model parameter representing source memory for trustworthy characters, d_{Trust}. We can test this hypothesis by imposing onto the model depicted in Fig. 1 the restriction that $d_{\mathrm{Cheat}}=d_{\mathrm{Trust}}$. This restricted model can then be fitted to the data. If the fit of the restricted model is significantly worse than the fit of the model without this restriction (and if, at the level of the estimates, $d_{\mathrm{Cheat}}>d_{\mathrm{Trust}}$), then we would have to reject that $d_{\mathrm{Cheat}}=d_{\mathrm{Trust}}$ and to conclude that source memory was better for cheater faces than for faces associated with trustworthiness.

Although the main purpose of the present series of experiments was to test whether *source memory* is improved for individuals with a history of cheating relative to individuals encountered in other situations, there was one additional consideration for the first three experiments reported here. We thought it possible that cheating behavior might be remembered better simply because it violates social norms and can therefore be considered exceptional and thus more distinct and memorable (Schmidt, 1991) than trustworthy behavior which is supposed to be the norm. In order to test this hypothesis, we manipulated the degree to which the behavior descriptions were exceptional as opposed to ordinary. In other words, for each of the behavioral history types (cheating, irrelevant, trustworthy) we used both exceptional and ordinary behavior descriptions. If the degree to which cheating behavior is exceptional determined how memorable a description was, then we should observe better source memory for exceptional than for ordinary behavior descriptions regardless of whether they represent cheating or trustworthiness.

3. Experiment 1

3.1. Method

3.1.1. Participants

Participants were 57 female and 28 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. They were paid for participating. Their age ranged from 19 to 46 years (mean=24, S.D.=5.1).

3.1.2. Apparatus and materials

A total of 72 facial photographs of males (256 bit, 116×164 pixel grayscales) were randomly assigned to two sets of 36 photographs each (henceforth Sets 1 and 2). Descriptions typed below the photographs conveyed the behavioral history (cheating, irrelevant to the cheating–trustworthiness dimension, trustworthy) of the person shown. Within each of these tree types of descriptions, half were classified as exceptional and half were classified as ordinary. As in earlier studies (Mealey et al., 1996; Mehl & Buchner, 2008), the descriptions also included information about the depicted person's profession to indicate the person's social status. However, in order to keep the design as simple as possible, only low-status job titles were used. For instance, "K. S. is a used-car dealer. He regularly sells restored crash cars as supposedly accident-free and conceals serious defects from the customers." would convey an ordinary history of cheating, whereas "G. K. is a soldier. He constantly steals munitions and other equipment from the camp and sells it to criminals." would convey an exceptional history of cheating. "O. N. is a scaffolder. Presently, he works at a building site in southern Germany where several tenements and office buildings are to be built." would convey an ordinary behavior that is irrelevant to the cheating–trustworthiness dimension, whereas "H. T. is an assembly line worker. He is very interested in the Far East and, as a practicing Buddhist, he meditates everyday even in his lunch breaks." would convey exceptional behavior that is irrelevant to the cheating–trustworthiness dimension. "O. D. is a cheese monger. He strongly attends to sorting out old cheese immediately and allows his customers to try all his products." would convey ordinary trustworthy behavior, whereas "F. L. is a baker. He allows some homeless people from his neighborhood to have breakfast and, in the winter, to have some hot coffee for free." would convey exceptional trustworthy behavior. In German, all sentences were 20 words long, not including the two initial letters representing the person's fictitious names.

Information about the social status of the professions, the valence of the behavior descriptions and the degree to which the described behaviors can be considered exceptional was obtained in independent norming studies. In one norming study, participants (N=36) rated 200 job titles with respect to their social status using a scale ranging from 1 (*low status*) to 5 (*high status*). A total of 36 job titles with low ratings were chosen for the experiment (mean=1.88, S.D.=.33; this is close to the status ratings of the professions used in the present Experiment 4 and in Mehl and Buchner, 2008). A different group of participants (N=22) rated the valence of 72 behavior descriptions to make sure that instances of cheating, irrelevant and trustworthy behavior were perceived as negative, neutral and positive, respectively. Valence was assessed on a scale ranging from −3 (*negative*) to +3 (*positive*). Half of the descriptive sentences were thought to specify exceptional behavior, whereas the other half were supposed to represent ordinary behavior. Therefore, participants were also asked to rate the 72 statements with respect to the exceptionality of behavior, using a scale raging from −3 (*very exceptional*) to +3 (*very ordinary*). Finally, six sentences were selected for

Table 1
Properties of the behavior descriptions used in the present experiments

	History of cheating	Irrelevant information	History of trustworthiness	Average
Valence (Experiments 1, 2 and 3)				
Exceptional	−2.83 (0.37)	0.30 (0.61)	2.53 (0.75)	0.00 (0.57)
Ordinary	−2.16 (0.64)	0.21 (0.58)	2.11 (0.73)	0.30 (0.61)
Average	−2.50 (0.51)	0.25 (0.59)	2.32 (0.74)	
Degree to which behavior is exceptional (Experiments 1, 2 and 3)				
Exceptional	−1.76 (1.36)	−1.04 (1.45)	−1.11 (1.48)	−1.30 (1.43)
Ordinary	0.14 (1.60)	1.53 (1.19)	0.61 (1.30)	0.76 (1.36)
Average	−0.81 (1.48)	0.21 (1.33)	−0.22 (1.37)	
Valence (Experiment 4)				
High social status	−2.29 (.84)	0.17 (.35)	1.84 (.75)	−0.10 (.26)
Low social status	−2.41 (.68)	0.27 (.31)	1.67 (.80)	−0.16 (.23)
Average	−2.35 (.72)	0.22 (.28)	1.75 (.73)	
Social status (Experiment 4)				
High social status	4.06 (.40)	4.00 (.45)	3.96 (.48)	4.00 (.33)
Low social status	2.04 (.41)	1.62 (.45)	1.82 (.44)	1.83 (.40)
Average	3.05 (.24)	2.81 (.25)	2.89 (.28)	

Values represent sample means (means of the item; standard deviations in parentheses). Valence ratings ranged from −3 (*negative*) to +3 (*positive*). Ratings of the degree to which behavior can be considered exceptional ranged from −3 (*very exceptional*) to +3 (*very ordinary*). Status ratings ranged from 1 (*low status*) to 5 (*high status*).

each category of the 3 (cheating, irrelevant, trustworthy)-by-2 (exceptional, ordinary) design. The properties of these sets of sentences are documented in Table 1. Invariably, there are still small differences in the rated dimensions among the types of photographs. For instance, the absolute valence (that is, ignoring the minus sign) is slightly larger for cheating (2.50) than for trustworthy (2.32) descriptions. Note, however, that the valence difference between exceptional (2.68) and ordinary (2.14) descriptions is three times larger. To anticipate, source memory did not differ at all between faces with exceptional and ordinary descriptions. This precludes that source memory differences between faces with cheating and trustworthy descriptions could have been caused by their much smaller differences in absolute valence. Photographs and descriptions were combined randomly for each participant.

3.1.3. Procedure

Participants were tested individually. They were asked to rate the attractiveness of 36 (Set 1 or 2, counterbalanced across participants) facial photographs that were presented in random order during the exposition phase. Each trial started with a headline (“How attractive do you find this person?”) and a photograph. The behavior description was shown 2 s later, followed 4.5 s later by the attractiveness rating scale [ranging from 1 (*not attractive at all*) to 6 (*extremely attractive*)]. Participants rated the attractiveness using the computer mouse and then initiated the next trial.

As in Mehl and Buchner’s (2008) Experiment 3, the exposition phase was directly followed by the test phase. Here, participants saw a random sequence of 72 photographs, half of which had been presented before (Set 1 or 2, depending on the exposition-phase assignment) and half were new (Set 2 or 1). Each trial started with a headline (“How attractive do you find this person?”) and a photograph. The attractiveness rating scale appeared 1.5 s later. After the rating, a new headline appeared (“Is this face old or new?”), followed by an “old” and a “new” checkbox, one of which participants selected to indicate that they had seen a face during the exposition phase or not. Following an “old” judgment and a click on the *continue* button, checkboxes labeled “cheating,” “trustworthy” and “neither cheating nor trustworthy” appeared which participants used to judge the behavior that was used in the description accompanying that faces in the exposition phase. After selecting one of these checkboxes and then clicking the *continue* button the next trial began.

3.1.4. Design

The within-subject independent variables were behavioral history (cheating, irrelevant, trustworthy) and whether the behavioral description was exceptional or ordinary. The dependent measures were attractiveness ratings, old–new discrimination in terms of hits (given that there was only one set of new faces in the test phase, there was only one false alarm rate for all types of faces so that sensitivity measures would be redundant) and source judgments given an “old” judgment.

Given a sample size of N=85, α=.05 and an assumed average population correlation between the levels of the behavioral history repeated measures variable of ρ=.55 (estimated from pilot data), effects of size f =0.17 [that is, between small (f=0.10) and medium (f=0.25) effects as defined by Cohen, 1977] could be detected for this variable with a probability of $1-\beta$=.95. All power calculations were conducted using G•Power (Faul, Erdfelder, Lang, &

Fig. 2. Ratings of faces and memory measures as a function of the behavior descriptions. “High” and “Low” in the figure legend refer to the degree to which the behavior descriptions in Experiments 1–3 were exceptional and to the social status of the profession associated with the faces in Experiment 4. *Ratings of Photographs*: Mean exposition-phase ratings of attractiveness (Experiment 1) and likeability (Experiments 2–4) on a scale from 1 (*not attractive/likeable at all*) to 6 (*extremely attractive/likeable*). Error bars represent the standard errors of the means. *Old–New Discrimination*: Test-phase old–new discrimination in terms of the mean number of hits (6 at most). The overall hit and false alarm rates were .77 and .11 in Experiment 1, .75 and .07 in Experiment 2, .68 and .12 in Experiment 3, and .76 and .13 in Experiment 4. Error bars represent the standard errors of the means. *Source Classification*: Mean number of correct classifications of faces associated with cheating, irrelevant information and trustworthiness as “cheating,” “neither cheating nor trustworthy” and “trustworthy,” respectively (12 at most). Error bars represent the standard errors of the means. *Source Memory Parameter d*: Parameter estimates for the source memory parameters for faces associated with a history of cheating (d_{Cheat}), for faces associated with irrelevant information ($d_{Irrelevant}$) and for faces associated with a history of trustworthiness (d_{Trust}). The parameters represent conditional probabilities of correct source identifications given correct old–new discriminations [combined D in Base Model 1 (see Table 2)=.67, .68, .56 and .63 in Experiments 1, 2, 3 and 4]. Error bars represent the .95 confidence intervals.

Buchner, 2007). A multivariate approach was used for all within-subject comparisons. In the present application, all multivariate test criteria correspond to the same (exact) F-statistic, which is reported. Partial η^2 is reported as a measure of the size of an effect. The level of α was set to .05, except for post hoc tests for which the significance level was Bonferroni–Holm corrected (Holm, 1979).

3.2. Results

3.2.1. Exposition-phase attractiveness ratings

A 3×2 MANOVA showed that the exposition-phase attractiveness ratings (Fig. 2) differed as a function of the behavioral history variable [$F(2, 83)=38.63$, $p<.001$, $\eta^2=.48$]. Post hoc contrasts showed that faces associated

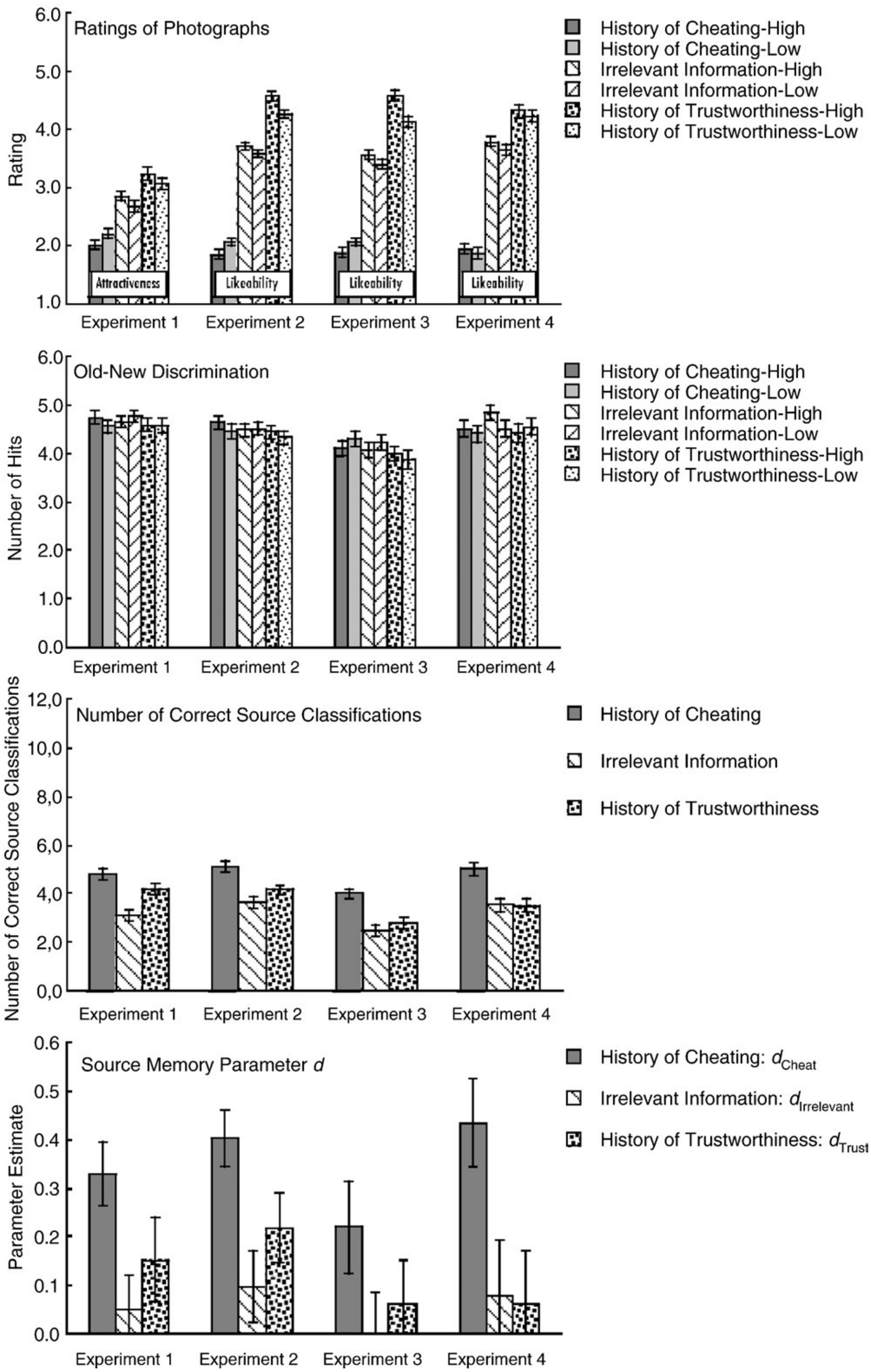

Table 2
Model-based results of source memory (see text for details)

Model test (parameter restriction)	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Base Model 1				
$D_{CheatHigh}$=$D_{CheatLow}$=$D_{IrrelevantHigh}$=$D_{IrrelevantLow}$= $D_{TrustHigh}$=$D_{TrustLow}$=D_{New} [a]	$G^2(9)$=4.51, p=.88	$G^2(9)$=9.31, p=.41	$G^2(9)$=14.40, p=.11	$G^2(9)$=8.06, p=.53
Additional restriction on Base Model 1				
$d_{CheatHigh}$=$d_{CheatLow}$, $d_{IrrelevantHigh}$=$d_{IrrelevantLow}$, $d_{TrustHigh}$=$d_{TrustLow}$ [a]	$\Delta G^2(3)$=5.71, p=.12	$\Delta G^2(3)$=2.66, p=.45	$\Delta G^2(3)$=0.33, p=.95	$\Delta G^2(3)$=1.05, p=.79
Base Model 2				
$D_{CheatHigh}$=$D_{CheatLow}$=$D_{IrrelevantHigh}$=$D_{IrrelevantLow}$= $D_{TrustHigh}$=$D_{TrustLow}$=D_{New} $d_{CheatHigh}$=$d_{CheatLow}$, $d_{IrrelevantHigh}$=$d_{IrrelevantLow}$, $d_{TrustHigh}$=$d_{TrustLow}$ [a]	$G^2(12)$=10.22, p=.60	$G^2(12)$=11.97, p=.45	$G^2(12)$=14.72, p=.26	$G^2(12)$=9.10, p=.69
Additional restriction on Base Model 2				
d_{Cheat}=d_{Trust}	$\Delta G^2(1)$=9.07, *p<.01*	$\Delta G^2(1)$=14.17, *p<.001*	$\Delta G^2(1)$=4.63, *p=.03*	$\Delta G^2(1)$=26.24, *p<.001*
Additional restriction on Base Model 2				
$d_{Irrelevant}$=d_{Trust}	$\Delta G^2(1)$=2.46, p=.12	$\Delta G^2(1)$=4.29, *p=.04*	$\Delta G^2(1)$=1.57, p=.21	$\Delta G^2(1)$=0.04, p=.84

The goodness-of-fit of the base models and the additional restrictions was tested using the goodness-of-fit statistic G^2, which is asymptotically χ^2 distributed with degrees of freedom indicated in parentheses. *P* values smaller than .05 (printed in italics) indicate that the implemented restrictions are not compatible with the data, as a result of which the hypothesis implied by the parameter restriction must be rejected.

[a] "High" and "low" in the indices refer to the degree to which the behavior descriptions in Experiments 1–3 were exceptional (i.e., exceptional vs. ordinary) and to the social status of the profession associated with the faces in Experiment 4.

with cheating were less attractive than the other two types of faces [F(1, 84)=77.62, p<.001, η^2=.48], and faces associated with irrelevant information were less attractive than faces associated with trustworthiness [F(1, 84)=34.83, p<.001, η^2=.29]. Faces associated with exceptional and ordinary behavior were equally attractive [F(1, 84)=2.27, p>.13, η^2=.03], but the interaction between both variables was significant [F(2, 83)=8.53, p<.001, η^2=.17], reflecting the fact that the attractiveness ratings were somewhat more negative for exceptional than for ordinary cheating behavior and somewhat more positive for exceptional than for ordinary irrelevant or trustworthy behavior. These results show that the behavior descriptions were processed, a necessary precondition for analyzing subsequent effects of the descriptions.

3.2.2. Old–new discrimination

A 3×2 MANOVA showed that behavioral history did not affect old–new discrimination in terms of the number of hits (Fig. 2) [F(2, 83)=0.76, p=.47, η^2=.02]. Faces associated with exceptional and ordinary behavior were recognized equally often [F(1, 84)=0.10, p=.76, η^2<.01]. The interaction between both variables was not significant [F(2, 83)=1.17, p=.32, η^2=.03].

3.2.3. Source memory

Given these results, it is now interesting to look at participants' source memory, that is, their memory for the behavior context in which a face was encountered. In a first step, it may be helpful to look at the raw number of correct source classifications for faces associated with cheating, irrelevant information or trustworthiness (Fig. 2). These classifications were more accurate for faces associated with cheating than for other types of faces [F(1, 84)=13.42, p<.001, η^2=.14]. Unfortunately, as we have seen in the Measuring Source Memory section, correct cheater classifications can be arrived at based on many different types of processes. We thus need to decompose classification performance into the processes involved in order to see whether it is really source memory that is better for cheaters than for other types of faces. For that purpose, participants' responses to all faces were analyzed using the multinomial model illustrated in Fig. 1. However, for an analysis of the present data we need two sets of the first three model trees, one set of three trees for *exceptional* cheating, irrelevant and trustworthy behavior, and another set of three trees for *ordinary* cheating, irrelevant and trustworthy behavior. Correspondingly, there are now also two sets of the parameters that occur in these trees. For instance, there is one parameter representing source memory for exceptional cheating behavior, $d_{CheatHigh}$, and one parameter representing source memory for ordinary cheating behavior, $d_{CheatLow}$ (*High* and *Low* in the index represent the degree to which the behavior was considered exceptional).

In order to simplify our analysis, we began with a base model (henceforth Base Model 1) that builds on the fact that the old–new discrimination judgments did not differ as a function of any of the independent variables. We thus decided for Base Model 1 to set all parameters to be equal that represent the probability of recognizing a face from the exposition phase as old. Based on the well-known mirror effect (Glanzer, Adams, Iverson, & Kisok, 1993), we also set the parameter representing the probability of detecting new faces as new to be equal to the face recognition parameters so that Base Model 1 is characterized by the general restriction that $D_{CheatHigh}$=$D_{CheatLow}$=$D_{IrrelevantHigh}$=$D_{IrrelevantLow}$=$D_{TrustHigh}$=$D_{TrustLow}$=D_{New}. These restrictions imply the assumption that the recognition of the faces was independent of whether they were presented as faces of cheaters, of persons with a description irrelevant to the cheating–trustworthy dimension or of trustworthy persons. This assumption was justified because, as can be seen in Table 2, Base Model 1 fits the data extremely well. Next, we tested whether the source memory parameters differed as a function of whether the behavior descriptions were exceptional or ordinary. This hypothesis can be implemented by imposing, on Base Model 1, the restrictions that $d_{\mathrm{CheatHigh}}=d_{\mathrm{CheatLow}}$, $d_{\mathrm{IrrelevantHigh}}=d_{\mathrm{IrrelevantLow}}$, and $d_{\mathrm{TrustHigh}}=d_{\mathrm{TrustLow}}$. These restrictions generate 3 degrees of freedom (*df*) in addition to the 9 *df* of Base Model 1. The corresponding *increase* in the model misfit as expressed in the goodness-of-fit statistic, ΔG^2, is asymptotically χ^2 distributed with 3 *df*. The second row of Table 2 shows that these restrictions were also compatible with the data. We thus conclude that source memory did not differ as a function of whether the behavior descriptions were exceptional or ordinary. This result was replicated in all subsequent experiments. We therefore combined the set of restrictions applied so far into our Base Model 2, which, as expected, also fitted the data well (see the third row of Table 2). The remaining hypothesis tests will be tested using Base Model 2.

First, we tested the central hypothesis that source memory is better for faces encountered in a cheating context than for faces associated with a context of trustworthiness. Descriptively this appears to be the case (Fig. 2). The null hypothesis of no such difference is tested by imposing, on Base Model 2, the restriction that $d_{\mathrm{Cheat}}=d_{\mathrm{Trust}}$, where d_{Cheat} and d_{Trust} represent source memory combined for exceptional and ordinary descriptions of cheating and trustworthy behavior, respectively. The restriction generates one additional degree of freedom. The corresponding increase in the goodness-of-fit statistic G^2 over that of Base Model 2 is asymptotically χ^2 distributed with 1 *df*. The fourth row of Table 2 shows that the restriction was clearly not compatible with the data. We must reject the hypothesis of no difference in source memory between cheater and trustworthy contexts, and instead conclude that source memory for faces of cheaters is significantly better than source memory for faces of trustworthy persons.

Second, we tested whether the descriptive difference between parameters $d_{\mathrm{Irrelevant}}$ and d_{Trust} was statistically significant. The statistical test of this hypothesis is parallel to the one just described. The result displayed in the fifth row of Table 2 shows that this difference was not statistically significant. We thus conclude that source memory for faces associated with trustworthy behavior does not differ from source memory for faces associated with irrelevant behavior.

3.2.4. Test-phase attractiveness ratings

An analysis of the test-phase attractiveness ratings (Table 3) showed a significant main effect of behavioral history [$F(2,83)=3.89$, $p=.02$, $\eta^2=.09$]. Cheater faces were less attractive than other faces [$F(1,84)=5.65$, $p=.02$, $\eta^2=.06$], whereas attractiveness did not differ between faces associated with irrelevant and trustworthy behavior [$F(1,84)=1.74$, $p=.19$, $\eta^2=.02$]. Faces associated with exceptional and ordinary behavior were equally attractive [$F(1,84)=0.07$, $p=.80$, $\eta^2<.01$]. There was no interaction between the two variables [$F(2,83)=2.97$, $p=.06$, $\eta^2=.07$].

Table 3
Test-phase attractiveness ratings (Experiment 1) and likeability ratings (Experiments 2, 3 and 4)

	History of cheating	Irrelevant information	History of trustworthiness
Experiment 1 (Attractiveness)			
Exceptional	2.44 (0.08)	2.60 (0.08)	2.65 (0.08)
Ordinary	2.56 (0.08)	2.56 (0.09)	2.61 (0.09)
Experiment 2 (Likeability)			
Exceptional	2.99 (0.06)	3.18 (0.06)	3.25 (0.06)
Ordinary	3.01 (0.07)	3.21 (0.06)	3.25 (0.07)
Experiment 3 (Likeability)			
Exceptional	3.07 (0.06)	3.22 (0.05)	3.11 (0.07)
Ordinary	3.06 (0.07	3.23 (0.06)	3.22 (0.07)
Experiment 4 (Likeability)			
High social status	3.05 (0.07)	3.44 (0.08)	3.27 (0.07)
Low social status	3.08 (0.08)	3.26 (0.08)	3.36 (0.08)

Values represent sample means (standard errors of the means in parentheses).

3.3. Discussion

The results of Experiment 1 replicate those of earlier experiments (Barclay & Lalumière, 2006; Mehl & Buchner, 2008) in showing that old–new discrimination of faces does not differ as a function of whether the faces were originally associated with cheating, behavior irrelevant to the cheating–trustworthiness dimension or trustworthiness. Contrary to the reasoning of Mealey et al. (1996), this is not inconsistent with the assumption that humans are equipped with special cognitive mechanisms supporting cheater avoidance in social exchange situations, because old–new discrimination per se cannot be useful when it comes to avoiding cheaters in future interactions. Given no difference in old–new discrimination, better memory for the source or context in which a face was encountered may well help to avoid cheaters. Compatible with this assumption, source memory was better for cheater faces than for faces associated with a history of trustworthiness.

In the test phase, cheater faces were rated less attractive than other faces. This may reflect participants' memory for the behavior associated with the face which translates into a generally negative reaction toward that person, just like during the exposition phase, but it could also reflect an unconscious negative reaction. Either way such a negative reaction might help to avoid costly social exchanges with cheaters, but note that the effect was rather small ($\eta^2=.09$).

Experiment 2 was designed as a conceptual replication of Experiment 1, but with likeability ratings instead of attractiveness ratings. Intuitively at least, likeability seems more closely related to the cheating–trustworthiness dimension than attractiveness. Performing likeability ratings instead of attractiveness ratings during exposition could thus lead to better integration of the face and the behavioral history information, making it easier to find effects of the behavioral history manipulation on all memory measures, perhaps even on the old–new discrimination performance. If old–new discrimination performance would still not vary as a function of the behavioral history variable, then this would be an even stronger evidence against the original interpretation of the Mealey et al. (1996) data and would confirm previous failures to find such an effect (Barclay & Lalumière, 2006; Mehl & Buchner, 2008). Better integration of face and behavior information should be indicated by a larger effect of the behavioral history variable on the exposition-phase likeability ratings here relative to Experiment 1.

4. Experiment 2

4.1. Method

Participants were 68 female and 28 male persons who were paid for participating. Most of them were students at the Heinrich-Heine-Universität Düsseldorf. Their ages ranged from 19 to 48 years (mean=25, S.D.=6.2). They had not participated in Experiment 1.

Apparatus and Materials, *Procedure* and *Design* were as in Experiment 1, except that likeability ratings replaced the attractiveness ratings. Given N=96, α=.05 and ρ=.55, effects of size f=.16 of the behavioral history variable could be detected with a probability of $1-\beta$=.95.

4.2. Results

4.2.1. Exposition-phase likeability ratings

A 3×2 MANOVA showed a main effect of behavioral history on the exposition-phase likeability ratings (Fig. 2) [F(2, 94)=277.35, p<.001, η^2=.86]. Faces associated with cheating were less likeable than the other two types of faces [F(1, 95)=548.54, p<.001, η^2=.85], and faces associated with irrelevant information were less likeable than faces associated with trustworthiness [F(1, 95)=163.71, p<.001, η^2=.63]. Faces associated with exceptional and ordinary behavior were equally likable [F(1, 95)=3.26, p=.07, η^2=.03]. The interaction between both variables was significant [F(2, 94)=23.85, p<.001, η^2=.34], reflecting the same trends as in Experiment 1.

4.2.2. Old–new discrimination

A 3×2 MANOVA showed that the number of hits (Fig. 2) did not differ as a function of the behavioral history variable [F(2, 94)=1.00, p=.37, η^2=.02]. Faces associated with exceptional and ordinary behavior were recognized equally often [F(1, 95)=1.19, p=.28, η^2=.01]. The interaction between both variables was not significant [F(2, 94)=0.46, p=.63, η^2=.01].

4.2.3. Source memory

The raw correct source classifications (Fig. 2) were more accurate for faces associated with cheating than for other types of faces [F(1, 95)=27.31, p<.001, η^2=.22]. Next, we need to decompose classification performance into the processes involved. Base Models 1 and 2 of the model illustrated in Fig. 1 fitted the data again very well (Table 2). For the central hypothesis test, the fourth row of Table 2 shows that the $d_{Cheat}=d_{Trust}$ restriction on Base Model 2 was incompatible with the data. Given the positive difference between the estimates of d_{Cheat} and of d_{Trust} (Fig. 2), we conclude that source memory was better for faces of cheaters than for faces of trustworthy persons. The fifth row of Table 2 shows that the descriptive difference between parameters $d_{Irrelevant}$ and d_{Trust} was statistically significant, although barely so. We may thus conclude that source memory for faces associated with trustworthy behavior was not quite as bad as source memory for faces associated with irrelevant behavior.

4.2.4. Test-phase likeability ratings

A supplementary analysis of the test-phase likeability ratings (Table 3) showed a significant main effect of behavioral history [F(2,94)=12.79, p<.01, η^2=.21]. Cheater faces were less likeable than other faces [F(1,95)=24.83, p<.01, η^2=.21], whereas likeability did not differ between faces associated with irrelevant and trustworthy behavior [F(1,95)=1.56, p=.21, η^2=.02]. Faces associated with exceptional and ordinary behavior were equally likable [F(1,95)=0.14, p=.71, η^2<.01]. There was no interaction between both variables [F(2,94)=0.06, p=.94, η^2<.01].

4.3. Discussion

The results of Experiment 2 closely replicate those of Experiment 1, but with likeability instead of attractiveness ratings. Most importantly, source memory was again best for faces associated with cheating. The effect of the behavior descriptions was much larger on the likeability ratings than on the attractiveness ratings of Experiment 1, suggesting that the likeability ratings involve better integration of the face and the behavioral history information. This should make it easier to find any effects of the behavioral history manipulation on the memory measures. Indeed, Fig. 2 shows that the source memory parameter d_{Cheat} was descriptively larger in Experiment 2 than in Experiment 1 (by 23%). Nevertheless, there was still no effect on the old–new discrimination, strongly confirming earlier failures to replicate the Mealey et al. (1996) findings (Barclay & Lalumière, 2006; Mehl & Buchner, 2008). Source memory, and not old–new discrimination, is better for faces of cheaters compared to other types of faces.

In the present Experiments 1 and 2, the retention interval between the exposition and the test phase was minimal, but this interval was 1 week long in the original study reported by Mealey et al. (1996) as well as in the experiments reported by Barclay and Lalumière (2006) and in Experiments 1 and 2 of Mehl and Buchner (2008). This raises the concern that the effect of the behavioral descriptions on source memory might perhaps vanish after a longer retention interval, so that the current findings would have limited implications. This suspicion is substantiated by the findings of Jacoby, Kelley, Brown, and Jasechko (1989). They showed that nonfamous names presented once in an experiment were mistakenly judged as famous 24 h later. On an immediate test, no such false fame occurred, presumably because the source of the names' familiarity could be recollected on the immediate test, but was unavailable 24 h later, as a consequence of which familiarity was misattributed to the current judgment dimension (fame). We therefore increased, in Experiment 3, the retention interval from a few minutes to 1 week.

5. Experiment 3

5.1. Method

Participants were 51 female and 21 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. Their ages ranged from 20 to 35 years (mean=23, S.D.=2.9). They had not participated in Experiment 1 or 2 and were paid for participating.

Apparatus and Materials, *Procedure* and *Design* were identical to those of Experiment 2, except that the retention interval was increased to 1 week. Given N=72, α=.05 and ρ=.55, effects of size f=.18 of the behavioral history variable could be detected with a probability of $1-\beta$=.95.

5.2. Results

5.2.1. Exposition-phase likeability ratings

A 3×2 MANOVA showed that the exposition-phase likeability ratings (Fig. 2) differed as a function of the behavioral history variable [$F(2, 70)$=155.70, p<.001, η^2=.82]. Post hoc contrasts showed that faces associated with cheating were rated significantly less likeable than the other two types of faces [$F(1, 71)$=274.31, p<.001, η^2=.79], and faces associated with irrelevant information were rated less likeable than faces associated with trustworthiness [$F(1, 71)$=195.18, p<.001, η^2=.73]. Faces associated with exceptional behavior were rated as more likeable than faces associated with ordinary behavior [$F(1, 71)$=13.95, p<.001, η^2=.16], which is different from the previous experiments. Parallel to Experiments 1 and 2, the interaction between both variables was significant [$F(2, 70)$=15.75, p<.001, η^2=.31], reflecting the same trends as in these experiments.

5.2.2. Old–new discrimination

A 3×2 MANOVA showed that the number of hits (Fig. 2) did not differ as a function of the behavioral history variable [$F(2, 70)$=2.73, p=.07, η^2=.07]. Faces associated with exceptional and ordinary behavior were recognized equally often [$F(1, 71)$=0.47, p=.50, η^2=.01]. The interaction between both variables was not significant [$F(2, 70)$=0.72, p=.49, η^2=.02].

5.2.3. Source memory

The raw correct source classifications (Fig. 2) were again more accurate for faces associated with cheating than for other types of faces [$F(1, 71)$=34.36, p<.001, η^2=.33]. Next, we need to decompose classification performance into the processes involved. Base Models 1 and 2 of the model illustrated in Fig. 1 fitted the data again very well (Table 2). For the central hypothesis test, the fourth row of Table 2 shows that the d_{Cheat}=d_{Trust} restriction on Base Model 2 was incompatible with the data. Given the positive difference between the estimates of d_{Cheat} and of d_{Trust} (Fig. 2), we conclude that source memory was better for faces of cheaters than for faces of trustworthy persons. Furthermore, the fifth row of Table 2 shows that the descriptive difference between parameters $d_{\text{Irrelevant}}$ and d_{Trust} was not statistically significant. We thus conclude that source memory did not differ between faces associated with trustworthy and irrelevant behavior.

5.2.4. Test-phase likeability ratings

A supplementary analysis of the test-phase likeability ratings (Table 3) showed a significant main effect of behavioral history [$F(2,70)$=3.93, p=.02, η^2=.10]. Cheater faces were less likeable than other faces [$F(1,71)$=5.67, p=.02, η^2=.07], whereas likeability ratings did not differ between faces associated with irrelevant and trustworthy behavior [$F(1,71)$=1.23, p=.27, η^2=.02]. Faces associated with exceptional and ordinary behavior were equally likable [$F(1,71)$=0.62, p=.43, η^2<.01]. There was no interaction between both variables [$F(2,70)$=0.90, p=.41, η^2=.03].

5.3. Discussion

Experiment 3 very closely replicated the findings of Experiments 1 and 2, thus demonstrating that the length of the retention interval does not affect the basic pattern of results, at least as long as the retention interval is 1 week or less.

Social status was manipulated in the study reported by Mealey et al. (1996) as well as in the experiments reported by Barclay and Lalumière (2006) and in Experiments 1 and 2 of Mehl and Buchner (2008). Whereas Mealey et al. found that social status unexpectedly modulated the effect of the behavioral descriptions on old–new discrimination, no such effect was found in any of the other experiments. It thus seemed important to test whether social status would modulate the effect of behavioral history on source memory. In order to keep the complexity of the design low, we dropped the exceptional–ordinary manipulation. This seemed justified given that this manipulation had no effect on any of the memory measures in Experiments 1, 2 and 3.

6. Experiment 4

6.1. Method

Participants were 31 female and 20 male persons, most of whom were students at the Heinrich-Heine-Universität Düsseldorf. Their ages ranged from 19 to 34 years (mean=23, S.D.=4.0). They had not participated in Experiment 1, 2 or 3 and were paid for participating.

Apparatus and Materials, Procedure and *Design* were identical to those of Experiment 2 with the following exceptions. The brief descriptions typed below the photographs were those used in Experiments 1 and 2 of Mehl and Buchner (2008). They conveyed the behavioral history (cheating, irrelevant, trustworthy) and the social status (low, high) of the person shown. Social status was conveyed through the profession of the person shown. Based on an independent norming study (N=24), professions with the lowest (20 out of 82) and highest (20 out of 82) ratings were selected and linked with cheating, irrelevant or trustworthy behaviors. An independent group (N=21) rated the valence of each behavior description. Finally, six sentences were selected for each of the categories of the 3 (behavioral history: cheating, irrelevant, trustworthy)-by-2 (social status: low, high) design. The properties of these sets of sentences are documented in Table 1. Given N=51, α=.05 and ρ=.55, effects of size f=.22 of the behavioral history variable could be detected with a probability of $1-\beta$=.95.

6.2. Results

6.2.1. Exposition-phase likeability ratings

A 3×2 MANOVA showed a main effect of behavioral history on the exposition-phase likeability ratings (Fig. 2) [F(2, 49)=103.07, p<.001, η^2=.81]. Cheater faces were less likeable than the other two types of faces [F(1, 50)=210.30, p<.001, η^2=.81], and faces associated with irrelevant information were less likeable than faces associated with trustworthiness [F(1, 50)=44.03, p<.001, η^2=.47]. Faces associated with high-status and low-status professions were equally likeable, but the effect just missed the preset significance level F(1, 50)=3.98, p=.052, η^2=.07]. The interaction between both variables was not significant [F(2, 49) =.31, p=.74, η^2=.01].

6.2.2. Old–new discrimination

A 3×2 MANOVA showed that the number of hits (Fig. 2) did not differ as a function of the behavioral history variable [F(2, 49)=1.63, p=.21, η^2=.06]. Faces associated with high- and low-status professions were recognized equally often [F(1, 50)=0.68, p=.42, η^2=.01]. The interaction between both variables was not significant [F(2, 49)=1.18, p=.32, η^2=.05].

6.2.3. Source memory

The raw correct source classifications (Fig. 2) were more accurate for faces associated with cheating than for other types of faces [F(1, 50)=19.31, p<.001, η^2=.28]. Next, we need to decompose classification performance into the processes involved. Base Models 1 and 2 of the model illustrated in Fig. 1 fitted the data again very well (Table 2). Source memory thus did not differ between low and high status descriptions. For the central hypothesis test, the fourth row of Table 2 shows that the $d_{\mathrm{Cheat}}=d_{\mathrm{Trust}}$ restriction on Base Model 2 was incompatible with the data. Given the positive difference between the estimates of d_{Cheat} and of d_{Trust} (Fig. 2), we conclude that source memory was better for faces of cheaters than for faces of trustworthy persons. The fifth row of Table 2 shows that there was no difference between parameters $d_{\mathrm{Irrelevant}}$ and d_{Trust}. We thus conclude that source memory did not differ between faces associated with trustworthy and irrelevant behavior.

6.2.4. Test-phase likeability ratings

A supplementary analysis of the test-phase likeability ratings (Table 3) showed a significant main effect of behavioral history [F(2,49)=12.59, p<.01, η^2=.34]. Cheater faces were rated less likeable than other faces [F(1,50)=24.76, p<.01, η^2=.33], whereas likeability ratings did not differ between faces associated with irrelevant and trustworthy behavior [F(1,50)=0.31, p=.58, η^2<.01]. There was no effect of social status on the test-phase likeability ratings [F(1,50)=0.81, p=.37, η^2=.02], and no interaction between the two variables [F(2,49)=2.47, p=.10, η^2=.09].

6.3. Discussion

Experiment 4 very closely replicated the findings of Experiments 1, 2 and 3. The effect of the behavioral history descriptions on source memory and test-phase likeability ratings did not interact with social status. An important aspect is that the behavioral descriptions used here were different from those used in Experiments 1, 2 and 3. Thus, the effects of the behavioral history variable on source memory were *not* due to particularities of the stimulus material.

7. General discussion

The present series of experiments yielded a consistent pattern of results. The behavioral histories associated with the faces had large effects on the exposition-phase attractiveness and likeability ratings, showing that participants processed the descriptions. Nevertheless, there was still no effect of the behavioral history variable on old–new discrimination in any of the present experiments. This result confirms earlier failures to replicate the Mealey et al. (1996) findings (Barclay & Lalumière, 2006; Mehl & Buchner, 2008) that old–new discrimination was better for cheaters than for other characters (although curiously only for their low-status behavior descriptions). Whatever caused this pattern of data, we may now safely conclude that the original Mealey et al. findings were very likely due to variables that are unrelated to the processing of cheater as opposed to other information.

However, these failures to replicate the Mealey et al. (1996) findings must not be counted as evidence against the idea that human memory is particularly good when it comes to avoiding cheaters in future interactions, let alone as evidence against social contract theory. This is so because avoiding cheaters is not possible only on the basis of knowing that one has seen a face before. Interestingly, given the preference often exhibited towards stimuli that are familiar but for which source information is unavailable (Bornstein, 1989; Zajonc, 1968), better recognition memory for cheaters (i.e., higher familiarity of cheaters) in the absence of source information might even imply a higher risk of being exploited. In order to avoid cheaters, it is necessary to recollect the source, that is, the cheating context in which they were encountered. Consistent with these considerations, the present findings show that while there was no cheater advantage in terms of old–new discrimination, source memory for cheater faces was selectively better than source memory for other types of faces with attractiveness (Experiment 1) and likeability ratings (Experiments 2, 3 and 4) in the exposition phase, with short (Experiments 1, 2 and 4) and long (Experiment 3) retention intervals, for both low-status and high-status characters (Experiment 4) as well as for both exceptional and ordinary behavior descriptions (Experiments 1, 2 and 3).

The latter finding is important in that it shows that the cheater effect cannot be reduced to an effect of the exceptionality of the behavior described. A priori, this seemed a serious possibility given that it has been previously observed that rarity within an experiment can modulate memory for cheaters (Barclay, 2008). Based on this finding, one could have argued that behavior that is in accordance with moral standards prevalent in the society is likely to be more common than behavior that violates social norms. Cheating behavior would thus automatically be considered exceptional behavior and as such could be more distinct and hence more memorable (Schmidt, 1991) than trustworthy behavior which is supposed to be the norm. The absence of an effect of the exceptionality of the behavior in Experiments 1, 2 and 3 shows that this is no viable explanation of the source memory advantage for cheaters observed in the present experiments.

Having excluded this alternative explanation for the cheater advantage in source memory, the present results can be considered indirect evidence in favor of social contract theory (Cosmides, 1989; Cosmides & Tooby, 1989; Tooby & Cosmides, 2005). According to this theory, social cooperation is so important to human evolution that special brain mechanisms have evolved that are functionally specialized to deal with it. For example, individuals have to detect the violation of social contract rules in order to engage successfully in social exchange. In addition, the brain mechanisms would also have to incorporate algorithms that "store information about the history of one's past exchanges with other individuals" (Cosmides & Tooby, 1992, p. 177).

The ability to recognize cheaters for what they are can be considered a prerequisite for the evolution and maintenance of social cooperation. This is so because social exchange cannot evolve in a species or be stably sustained in a social group unless individuals refuse to cooperate with individuals who have cheated on them in past encounters (direct reciprocity; Axelrod & Hamilton, 1981; Trivers, 1971) or refuse to cooperate with individuals who are known to have cheated in interactions with third parties (indirect reciprocity; Nowak & Sigmund, 2005), or unless cheating is punished directly (altruistic punishment; Fehr & Gächter, 2002). To serve these purposes, the cheater-detection module would have to include mechanisms that support the recall of individuals who cheated. The consistently better source memory for cheaters than for other types of faces observed here is clearly consistent with such an assumption.

It is less clear what we should make of the difference in source memory for trustworthy and irrelevant behavior descriptions. On the one hand, the assumption of a "cheater-detection module" alone does not allow us to predict any differences in source memory for trustworthy and irrelevant behavior descriptions. One could thus argue that, in fact, there was no such difference because the difference between d_{Trust} and $d_{\text{Irrelevant}}$ was statistically significant only in Experiment 2 (Table 2). On the other hand, one could take the position that there really was a source memory difference between trustworthy and irrelevant behavior descriptions, but that this difference was so small that it is difficult to detect. This fits with the fact that the experiment in which the difference between d_{Trust} and $d_{\text{Irrelevant}}$ was statistically significant was also the one with the largest number of participants and, hence, the highest statistical power. Also, when the data from all four experiments were combined, the difference between d_{Trust} and $d_{\text{Irrelevant}}$ was statistically significant [$G^2(1)$=6.81, p<.01], although the effect was much smaller than the difference between d_{Cheat} and d_{Trust} [$G^2(1)$=42.26]. Finally, the argument draws support from the consideration that knowing whom one can trust should also be of some value for future interactions, albeit it may be less important than knowing who cheats. We tend to favor the second of these options, but clearly the present data do not allow us to draw a firm conclusion about this issue.

In more general terms, the present data are compatible with a functional view of memory (Nairne & Pandeirada, 2008; Nairne, Pandeirada, & Thompson, 2008; Nairne, Thompson, & Pandeirada, 2007) according to which it may be particularly fruitful to analyze human memory performance with respect to past (and present) fitness advantages. Avoiding cheaters may be seen as an adaptive problem, which a cheater-specific source memory advantage may help to solve.

References

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

Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. *Cognition, 107*, 817−828.

Barclay, P., & Lalumière, M. L. (2006). Do people differentially remember cheaters? *Human Nature, 17*, 98−113.

Batchelder, W. H., Hu, X., & Riefer, D. M. (1994). Analysis of a model for source monitoring. In G. H. Fischer, & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 51−65). New York: Springer.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*, 548−564.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 22, 197−215.

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition, 28*, 480−500.

Bell, R., Buchner, A., & Mund, I. (2008). Age-related differences in irrelevant-speech effects. *Psychology and Aging, 23*, 377−391.

Bernstein, F. (1925). Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen [Summarizing considerations on the inheritable blood structures of mankind]. *Zeitschrift für Abstammungs- und Vererbungslehre, 37*, 237−270.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin, 106*, 265−289.

Burnstein, E. (2005). Altruism and genetic relatedness. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 528−551). Hoboken, NJ: John Wiley & Sons Inc.

Buss, D. M. (2004). *Evolutionary psychology: The new science of the mind.* Boston, MA: Allyn and Bacon.

Cartwright, J. (2000). *Evolution and human behavior: Darwinian perspectives on human nature.* Cambridge, MA: The MIT Press.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. (Rev. ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture: II. Case study: A computational theory of social exchange. *Ethology and Sociobiology, 10*, 51−97.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163−228). New York, NY: Oxford University Press.

Cummins, D. (2005). Dominance, status, and social hierarchies. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 676−697). Hoboken, NJ: John Wiley & Sons Inc.

Dodson, C. S., & Shimamura, A. P. (2000). Differential effects of cue dependency on item and source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1023−1044.

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology, 41*, 369−387.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G∗Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175−191.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137−140.

Gaulin, S. J. C., & McBurney, D. H. (2001). *Psychology: An evolutionary approach.* Upper Saddle River, NJ: Prentice Hall.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kisok, K. (1993). The regularities of recognition memory. *Psychological Review, 100*, 546−567.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65−70.

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika, 59*, 21−47.

Jacoby, L. L., Kelley, C. M., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology, 56*, 326−338.

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology, 17*, 119−128.

Mehl, B., & Buchner, A. (2008). No enhanced memory for faces of cheaters. *Evolution & Human Behavior, 29*, 35−41.

Murnane, K., & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition, 24*, 417−428.

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive memory: Remembering with a stone-age brain. *Current Directions in Psychological Science, 17*, 239−243.

Nairne, J. S., Pandeirada, J. N. S., & Thompson, S. R. (2008). Adaptive memory: The comparative value of survival processing. *Psychological Science, 19*, 176−180.

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 263−273.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*, 1291−1298.

Palmer, J. A., & Palmer, L. K. (2002). *Evolutionary psychology: The ultimate origins of human behavior.* Needham Heights, MA: Allyn and Bacon.

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 680−693.

Rothkegel, R. (1999). AppleTree: A multinomial processing tree modeling program for Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 31*, 696−700.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*, 523−542.

Simons, J. S., Verfaellie, M., Galton, C. J., Miller, B. L., Hodges, J. R., & Graham, K. S. (2002). Recollection-based memory in frontotemporal dementia: Implications for theories of long- term memory. *Brain: A Journal of Neurology, 125*, 2523−2536.

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 631−651.

Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods, 39*, 267−273.

Tooby, J., & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 5−67). Hoboken, NJ: John Wiley & Sons.

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

Zajonc, R. B. (1968). Attitudinal Effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1−27.