

How Surveillance Begets Perceptions of Dishonesty: The Case of the Counterfactual Sinner

Dale T. Miller
Stanford University

Penny S. Visser
University of Chicago

Brian D. Staub
Princeton University

Three studies support the hypothesis that observers' impressions of actors reflect not only what actors do but also what they can easily be imagined doing. Participants in Studies 1 and 2 observed a 10-year-old boy take a math test in a context in which the incentive to cheat and the constraints against cheating varied. When the incentive to cheat was high but the likelihood of getting caught was also high, observers perceived a target who resisted the temptation to cheat as less honest than the average boy. This effect was not found when the incentive to cheat was low, which suggests that its occurrence under high temptation resulted from observers in that condition generating the counterfactual thought that the target would have cheated had the likelihood of detection been low. Study 3 further supported the link between spontaneous counterfactual thought and inferences of dishonesty. The implications of the counterfactual correspondence bias are discussed.

Keywords: counterfactual thought, person perception, surveillance, trust

We estimate a man by imagining what he would do in various situations. (Mead, 1934, p. 105)

People virtually never encounter someone without expectations of how he or she will behave. Moreover, the expectations that people bring to their interpersonal encounters, whether they are based on beliefs about the situation or about the target, are notoriously powerful (Jones, 1990; Miller & Turnbull, 1986; Trope, 1986). They guide every aspect of the person perception process, including which features of the actor's behavior are attended to, how the actor's behavior is identified or categorized, and how the actor's behavior is explained (Olson, Roese, & Zanna, 1996).

Beliefs or naive theories guide more than the precomputed expectancies that observers bring with them to the interpersonal encounter, however. They also guide the postcomputed expectancies (counterfactual norms) that observers generate during the encounter (Kahneman & Miller, 1986). To illustrate the distinction between precomputed and postcomputed expectancies, consider how each might

contribute to feelings of disappointment in an unsuccessful job applicant. To begin with, the applicant can be expected to experience disappointment to the extent that her beliefs led her to feel that she would be successful. Even if the applicant expected to be unsuccessful, however, she still can be expected to experience disappointment to the extent that her beliefs led her to feel that she came very close to being successful, despite her initial pessimism. Her disappointment in the first instance derives from the contrast between her outcome and a precomputed expectancy, and in the second instance it derives from the contrast between her outcome and a postcomputed expectancy or counterfactual construction. Postcomputed expectancies, like their precomputed counterparts, powerfully influence perceptual and affective reactions to the behavior of others, including the explanation provided for the actor's behavior, the degree of blame assigned to the actor for the outcomes caused by his or her behavior, and the degree of sympathy or hostility evoked by the actor (Roese, 1997).

The present research seeks to show that the postcomputed or counterfactual thoughts evoked by a target's behavior affect even the most central aspect of the person perception process: trait inference. That observers insufficiently weight information about situational forces when interpreting another's behavior is virtually axiomatic in social psychology (Gilbert, 1998; Ross, 1977). As Heider (1958) memorably put it, "Behavior engulfs the field" (p. 54). We contend that counterfactual behavior (images of what might have been) can also engulf the field and lead to unwarranted person inferences. That is, observers draw unwarranted dispositional inferences not only from actual behavior that could be explained by situational inducements but also from imagined be-

Dale T. Miller, Graduate School of Business, Stanford University; Penny S. Visser, Department of Psychology, University of Chicago; Brian D. Staub, Department of Psychology, Princeton University.

This research was supported by National Institute of Mental Health Research Grant MH44069. We thank Lisa Pugh, Sarah Townsend, and Joe Magee for their assistance with the studies.

Correspondence concerning this article should be addressed to Dale T. Miller, Graduate School of Business, 518 Memorial Way, Stanford University, Stanford, CA 94305-5015. E-mail: dtmiller@psych.stanford.edu

havior that situational inducements bring to mind. For example, the inference that a person is dishonest may not require that the person be observed acting dishonestly; it may be sufficient that the person be observed in a situation that evokes images of another situation in which he or she likely would have acted dishonestly.

The Counterfactual Correspondence Bias

The correspondence bias and what we term the counterfactual correspondence bias both stem from people's tendency to treat nondiagnostic behavioral information as though it were diagnostic. Observers commit the correspondence bias when they lose sight of the fact that the information they have about the target suggests that he or she acted as the average person would in that situation and therefore should not be taken as evidence that he or she is nonaverage. Observers commit the counterfactual correspondence bias when they lose sight of the fact that the (counterfactual) information they have about the target was generated by their assumptions about the average person and therefore should not be taken as evidence that the target is nonaverage. The correspondence bias reflects people's inclination to infer dispositions from behavior that actors actually engaged in, but under conditions in which they could not have done otherwise. The counterfactual correspondence bias reflects people's inclination to infer dispositions from behavior that actors did not actually engage in but very well might have had the situation been different.

To illustrate the counterfactual correspondence bias at work, imagine watching a 10-year-old boy in a test situation in which the incentive to cheat is high but so are the chances of getting caught. Imagine further that you believe (a) that the average 10-year-old boy would cheat in this situation if he did not think he would get caught and (b) that virtually no 10-year-old boy would cheat in this situation if he thought he would get caught. How might you respond to the fact that this particular boy resisted cheating? Given that you did not expect him to cheat under these circumstances, his behavior should not surprise you. That is, it should not evoke contrasting images of different (more likely) behavior in the observed situation. What it might very well evoke, however, are thoughts of different (more likely) behaviors in the counterfactual situation in which the boy did not fear detection. You might, for instance, find yourself entertaining such thoughts as, "He's only behaving himself because he does not want to get caught." The consequence of having counterfactual thoughts of dishonest behavior, we propose, is that you will draw the inference that the target is inclined to dishonesty. In effect, you will treat the counterfactual evidence you have as though it were factual evidence.

The three studies that follow were designed to demonstrate the counterfactual correspondence bias as well as to probe the processes underlying it. All three studies use an experimental setting modeled after the example described above.

Study 1

Study 1 sought to demonstrate that observers' thoughts of what a person might have done had the situation been different can affect the trait inferences they draw about the person. Participants were shown videotapes of two targets in a situation in which there was a strong incentive to act dishonestly. For one target, a powerful situational constraint against acting dishonestly was present

(a video surveillance camera); for the other, there was no such situational constraint. The target resisted the temptation to act dishonestly in both situations.

The design permits the test of two hypotheses. The first is that the target who resists acting dishonestly under low constraint will be judged more honest than the target who resists acting dishonestly under high constraint. This hypothesis was first tested by Strickland (1958) in his classic study on surveillance and trust and derives from the fact that the cause of the former target's behavior is less ambiguous than is the cause of the latter target's behavior (Kelley, 1971). Specifically, although there are two plausible explanations for the constrained target's behavior (i.e., dispositional honesty or external constraints against behaving dishonestly), there is only one explanation for the unconstrained target's behavior (i.e., dispositional honesty).

According to Strickland's (1958) and Kelley's (1971) analyses, the greater honesty attributed to the unconstrained than to the constrained target (the Strickland effect) reflects the fact that something positive has been inferred about the former target in this regard, whereas nothing has been inferred about the latter target. Our analysis of the counterfactual correspondence bias, however, suggests that not all the inferential action that occurs in this situation pertains to the unconstrained target. It is true that observers will draw positive inferences about the unconstrained target, but they also will draw negative inferences about the constrained target. Thus, our second hypothesis predicts that observers will perceive a constrained target who resists acting dishonestly not only as less honest than an unconstrained target but also as less honest than the average person.

Method

Participants

Forty-four undergraduates participated individually in partial fulfillment of a course requirement. The study was advertised as being concerned with nonverbal communication and body posture.

Procedure

When participants arrived at the laboratory they were seated and given the following description of the study.

The current study is concerned with nonverbal behavior. Specifically, we are interested in identifying nonverbal cues and body posture that may be associated with cheating behaviors. In order to examine this, we brought schoolchildren (ages 6–18) into the lab and, using a hidden camera, videotaped them completing a set of 16 math problems. The correct answers to these math problems were placed on the board directly behind the children. The children were told that they were going to receive \$1 for every correct answer they were able to get in the time allowed. Intentionally, we made the time limit short so as not to allow the children enough time to successfully complete all the questions.

We randomly told half of the children about the hidden camera (this was accomplished by simply telling every other child who entered the lab about the camera). As a result, the child was only aware of being videotaped if the experimenter directly told him. We now have to review these tapes to analyze the behavior of the children. In order to do this effectively we need to have a large number of blind reviewers (i.e., people who do not know anything about the children in the study). As a result, we are utilizing students in the subject pool to review the videotapes for us. This will be your "role" in the experiment.

What you are about to do is simply watch videotapes of the children completing the math problems. Specifically, you will be watching 10–11 year olds because of the high occurrence of cheating discovered in that age group. You will watch 3 such tapes (out of hundreds of possible tapes), lasting approximately 2 min. If time allows, we will show you more tapes of other children. Please watch the tapes carefully and mentally note any behaviors you see—especially those that may be associated with cheating traits. We will not show you actual acts of cheating. Rather we are interested in your ability to pick up behavior that may be diagnostic of cheating (very subtle behaviors that may, e.g., help teachers better monitor cheating behavior).

We realize that this is a very difficult task. We only ask that you observe the children carefully and do the best you can on the questions that follow. Some of the questions will be opinion questions—answer these as honestly as you can. Other questions will actually have a correct answer (e.g., Which child was rated by his teacher as most curious?). We realize this may seem very difficult but please do the best you can—most people do surprisingly well.

On the top of the screen will be the child's name and whether or not he was aware of the hidden camera—please make sure you remember which students were aware of the camera and which ones were not.

Please read the background information now and look over the questionnaire. We want you to be familiar with the questions we plan to ask before we begin showing the tapes. After you have done this, we will start with the first tape.

Participants were then shown a page of background information (place of birth, parents, siblings, musical background, and math and reading levels, which were always average) on each of three boys, along with a photograph.

Stimulus tapes. Once they had completed reading the background information, participants were shown the videotapes without the sound. They had been told that they would see three tapes, but after the first two were finished the experimenter explained that he was running late and would not show them the third tape so that they could complete the next phase of the study. In counterbalanced order, one of the tapes displayed the label “aware of video” along with the child's name at the top, and one displayed the label “unaware of video” along with the child's name at the top. Each clip showed one of three 10-year-old male confederates actually working on math questions, with the answers prominently displayed on the board behind him. None of the confederates turned around to look at the answers on the board. The particular child who appeared as the target in each condition (aware, unaware, control) was counterbalanced.

Dependent measures. At the completion of the two clips, participants were asked to complete a questionnaire that asked about all three targets. The experimenter told the participants that he realized they had not had the opportunity to see the tape of the third boy but asked them to try their best to answer the questions posed. The questionnaire began with some general filler questions about the targets. It then asked participants to indicate on a 7-point scale how likely they thought it was that the monitored target would have cheated on the test had he not known about the video camera.

Participants were next told that we had compiled personality profiles of the boys on the basis of their schoolteachers' ratings of each boy on various characteristics and personality traits. The participants were then presented with a list of traits and asked to indicate which boy they thought had received the highest teacher rating for each trait. Embedded in the longer list were four key traits: (a) trustworthy, (b) dishonest, (c) deceitful, and (d) dependable. Next, participants were presented with a series of hypothetical positive and negative behaviors. For each behavior, participants were asked to indicate which of the three targets they thought was the one most likely to have performed the behavior. Embedded in the longer list were two key behaviors: (a) getting caught stealing money out of another student's book bag and (b) returning a lost wallet with money in it.

Following this measure, participants were presented with each possible pairing of the boys and in each case were asked to indicate which of the two they would choose to supervise in a future testing situation. Next, participants

were told that a typical student in this age group would be expected to correctly complete approximately 8 of the 16 math problems that were presented to the boys. With this in mind, participants were asked to estimate how many of the problems each of the targets had answered correctly.

Finally, to assess the effectiveness of the manipulations, we asked participants to indicate which one of the two boys they had observed knew he was being videotaped and which one did not, as well as how we decided who was told he was being videotaped and who was not. Once the participants had completed the questionnaire, they were probed for suspicion and debriefed.

Results

Manipulation Check

All participants correctly identified which actor was aware that he was being videotaped and correctly noted that the decision of whom to tell and whom not to tell was made randomly.

Evidence of the Strickland Effect

Did participants view a temptation-resisting target as more honest when he was unaware he was being monitored than when he was aware? We addressed this question in two ways.

Choice for supervision. To assess the perceived honesty of the two targets, we first compared the number of nominations each target received when participants were asked to indicate which of the two targets they would choose to supervise in the future if they could only supervise one of them. Replicating Strickland (1958), participants faced with the choice of supervising the monitored target or the unmonitored target overwhelmingly chose the former (75% vs. 25%), $\chi^2(1, N = 44) = 11.00, p = .001$.

Predictions about honesty. In pursuit of further evidence of the Strickland effect, we examined the participants' nominations of the target they thought was most likely (a) to be seen by his teachers as trustworthy, (b) to be seen by his teachers as dependable, and (c) to return a lost wallet with money in it. The responses to these questions were combined to create an *honesty index* for each target that could range from 0 to 3 (Cronbach's $\alpha = .74$). Consistent with predictions, the honesty index for the unmonitored target ($M = 1.86$) was significantly higher than that for the monitored target ($M = 0.73$), $t(43) = 3.61, p = .001$. The unmonitored target was also perceived to be more honest than the control target ($M = 0.41$), $t(43) = 5.61, p < .001$. The honesty indices for the monitored target and the control target did not differ significantly, $t(43) = 1.42$.¹

Evidence of the Counterfactual Correspondence Bias

Did participants view a temptation-resisting target who was aware he was being monitored as more dishonest than a control target about whom they had no behavioral information? We addressed this question in two ways.

¹ The nomination measure (i.e., the selection of one target from three) proved to be sensitive to differences between the most highly nominated target and the other two targets but not to differences between the latter two targets. The reason for this is that there were very few nominations of targets other than the unmonitored target on the honesty items and other than the monitored target on the dishonesty items. This means that differences between the monitored and control targets on the honesty index and between the unmonitored and control targets on the dishonesty index are not especially interesting.

Choice for supervision. To determine whether participants concluded that the monitored target was more dishonest than the average boy his age, we first compared the number of nominations he received from participants when they were asked to indicate whether they would choose to supervise him or the control target in the future. Consistent with predictions, faced with a choice of supervising the monitored target or the control target, participants chose the former (73% vs. 27%), $\chi^2(1, N = 44) = 9.09, p < .01$.

It appears that participants' previously reported belief that the monitored target warranted closer supervision than the unmonitored target reflected at least partly their belief that the evidence they had suggested that the latter was dishonest and not simply that the former was honest. Stated differently, the present results suggest that participants' decision about the target most in need of future supervision was not made without prejudice toward the monitored target and rests at least partly on their inference, in addition to their inference that the unmonitored target was honest, that the monitored target was dishonest.

Predictions about dishonesty. In pursuit of further evidence of a counterfactual correspondence bias, we examined the participants' nominations of the target they thought was most likely (a) to be seen by his teachers as dishonest, (b) to be seen by his teachers as deceitful, and (c) to be caught stealing money out of another student's book bag. The responses to these questions were combined to form a *dishonesty index* for each target that could range from 0 to 3 (Cronbach's $\alpha = .87$). Consistent with predictions, the dishonesty index for the monitored target ($M = 1.84$) was significantly higher than that for the unmonitored target ($M = 0.43$), $t(44) = 4.72, p < .001$. More important, the monitored target was also perceived to be more dishonest than the control target ($M = 0.73$), $t(44) = 3.25, p = .002$. Perceptions of the unmonitored and control targets did not differ significantly, $t(44) = 1.29$.

Ability Attribution

In addition to assessing participants' impressions of the targets' character, we also assessed their estimates of the targets' ability at math by examining the number of math problems they thought each of the three targets had correctly answered. (Recall that participants had been told that the average number of problems answered correctly by children of this age group was eight.) Participants estimated that the unmonitored target had scored significantly higher ($M = 8.70$) than the monitored target, ($M = 6.91$), $t(43) = 3.45, p = .001$. Furthermore, participants estimated that the unmonitored target had scored significantly higher than the test average, $t(43) = 2.03, p < .05$, and that the monitored target had scored significantly lower than both the test average, $t(43) = 2.81, p < .01$, and the control target ($M = 8.07$), $t(43) = 3.13, p < .01$.²

The finding that participants predicted that the unmonitored target had done better than average on the math test is reasonable in light of the fact that the target chose not to cheat on the test when he had the opportunity to do so. The finding that participants thought the monitored target had done worse than average is not intuitively obvious, but it is consistent with our reasoning. We assumed that, when observing the monitored target, observers would generate the thought that he would have cheated were he not monitored, a thought that, in addition to making the target seem like a cheater, would likely raise doubts about his ability. Consistent with this assumption, the more likely observers thought it was that the monitored target would have

cheated if not monitored, the worse they expected him to have done on the test ($r = -.48, p < .001$).

Counterfactual Thoughts and Target Evaluation

The monitored target. The prediction that participants would perceive the monitored target to be more dishonest than the control target was based on the assumption that participants, while viewing the monitored target, would generate the thought that he would have cheated (i.e., looked at the answers behind him) were it not for his knowledge that he was being monitored. Consistent with the proposed analysis, the more confident participants were that the monitored target would have cheated if he were not monitored, (a) the more likely they were to choose him rather than the control target for future supervision ($r = .43, p < .01$), (b) the higher the dishonesty score they assigned him ($r = .66, p < .001$), and (c) the fewer math problems they estimated he got correct on the test ($r = -.48, p < .001$).

The unmonitored target. The source of participants' counterfactual thoughts about the monitored target was presumed to be their *theory of the situation*, or, more specifically, their estimate of the inclination of the average 10-year-old boy to cheat in the experimental situation. As a consequence, we predicted that the strength of the cheating counterfactual would be correlated with the impressions that participants formed of the unmonitored target as well as those they formed of the monitored target. Specifically, the more likely a participant's theory suggested it was that the monitored target would have cheated if he had not been monitored, the more favorably impressed the participant should have been by the unmonitored target, who refrained from cheating despite his belief that he was not being monitored. Indeed, the more confident participants were that the monitored target would have cheated if not monitored, (a) the less likely they were to choose the unmonitored target over the control target for future supervision ($r = -.30, p = .05$) and (b) the higher was the honesty score they assigned to him ($r = .51, p < .001$). The strength of the counterfactual was unrelated to participants' estimates of the unmonitored target's math ability ($r = -.01$).

Discussion

Study 1 yields two principal findings. First, it provides a conceptual replication of Strickland (1958) by showing that participants viewed a child who resisted the temptation to cheat when he did not know that he was being monitored as more honest than they did one who knew he was being monitored. Second and more interesting, the study shows that participants viewed a temptation-resisting child who knew he was being monitored as more dishonest than they did one whom they did not observe. By this pattern of inference, participants revealed that they believed they had learned something—had acquired *person information*, to use Jones and Davis's (1965) term—about the monitored target as well as the unmonitored target.

It is important to note that the participants' inferences about the monitored target do not simply constitute another demonstration of the correspondence bias. Participants did not disregard situational information (i.e., the constraints on cheating) in favor of behavioral

² Supporting our assumption that participants viewed the control target as a proxy for the average 10-year-old boy, participants' estimates of the control target's performance were virtually identical to the average test score, $t(43) = 0.22, ns$.

information, for had they done so they would have inferred that the noncheating monitored target was less (not more) dishonest than average. Rather, their inferences constituted a counterfactual correspondence bias, for the inference drawn by participants corresponded not to actual behavior but to imagined or counterfactual behavior. In effect, the target was being held accountable not for what he did (i.e., not cheating) but for what he easily could be imagined doing (i.e., cheating) under different circumstances. Correlational analysis supported this reasoning by showing that the more confident participants were that the monitored target would have cheated were he not monitored, the more negatively they evaluated him.

Study 2

Study 2 sought further evidence for the claim that counterfactual thoughts of dishonesty can produce attributions of dishonesty. To do this, it manipulated the strength with which the target's behavior evoked counterfactual thoughts of dishonest behavior. More specifically, Study 2 manipulated the incentive the target had for cheating. We hypothesized that the greater the incentive was to cheat, (a) the more likely observers would be to assume that a monitored noncheating target would have cheated were he not constrained by the knowledge of the video camera and (b) the more likely observers would be to infer that the monitored noncheating target was dishonest.

Method

Participants

Sixty-seven undergraduates participated in this study in partial fulfillment of a course requirement.

Procedure

The cover story, stimulus materials, and dependent measures were highly similar to those used in Study 1. In fact, the procedure and measures used in the high incentive condition of Study 2 were identical to those of Study 1, with one exception. The description of the study omitted any reference to the base rate of cheating behavior in the target population.

For participants in the low incentive condition, the stimulus materials were modified in two additional ways. First, no mention was made of any financial incentive for good performance on the math problems. Second, whereas participants in the high incentive condition were told that we had intentionally made the time limit short so that the children would not have enough time to complete all the questions, participants in the low incentive condition were told that the children were provided ample time to complete all of the questions. In all other ways, the two conditions were identical.

Results

Manipulation Checks

All participants correctly identified which target knew he was being videotaped and that the decision of whom to tell and whom not to tell was random. In addition, confirming the effectiveness of the incentive manipulation, participants in the high incentive condition estimated that the likelihood that the monitored target would have cheated had he not known about the video camera was much higher ($M = 5.45$) than did participants in the low incentive condition ($M = 4.03$), $t(65) = 2.63, p = .01$.

Evidence of the Strickland Effect

To determine whether the unmonitored target was perceived to be more honest than the monitored target, we followed the same procedure as in Study 1. Specifically, we examined participants' choice for future supervision and their nominations for the most honest target. We expected that when the incentive to cheat was high, perceivers who observed the monitored target resist the temptation to cheat would be highly likely to generate the thought that he would succumb to this temptation in the absence of supervision. Conversely, when the incentive to cheat was low, we expected that perceivers would be much less likely to generate counterfactual thoughts of the monitored target behaving dishonestly in the absence of supervision. As a result, we predicted an interaction between surveillance status (unmonitored vs. monitored) and level of incentive (low vs. high) on participants' assessments of the targets.

Choice for supervision. To assess the perceived honesty of the two targets, we first compared the number of nominations each target received when participants were asked to indicate which of the two targets they would choose to supervise in the future if they could only supervise one of them. As expected, these nominations varied significantly across incentive levels, $\chi^2(1, N = 67) = 5.81, p < .02$. Consistent with the findings from Study 1, when the incentive to cheat was high, participants overwhelmingly opted to supervise the monitored target in a future session (76% vs. 24%), $\chi^2(1, N = 33) = 8.76, p < .01$. Participants in the low incentive condition, however, showed no preference for supervising the monitored target over the unmonitored target in a future session (47% vs. 53%), $\chi^2(1, N = 34) = 0.12$.

Predictions about honesty. We predicted an interaction between surveillance status (unmonitored vs. monitored) and level of incentive (low vs. high) on the honesty score participants assigned to the targets. To test this prediction, we conducted a repeated measures analysis of variance (ANOVA) on the honesty index (Cronbach's $\alpha = .75$), with surveillance status as a within-subject factor and incentive level as a between-subjects factor. As expected, the interaction was significant, $F(2, 64) = 5.58, p < .01$, which suggests that the impact of monitoring on perceivers' impressions of the target differed depending on the magnitude of the incentive to cheat. To determine the nature of the interaction, we examined the impact of monitoring on honesty ratings separately among participants in the high and low incentive conditions (see Figure 1).

In the high incentive condition, consistent with the results of Study 1, the honesty index for the unmonitored target ($M = 2.18$) was significantly higher than that for the monitored target ($M = 0.39$), $t(32) = 6.36, p < .001$. Also replicating Study 1, the unmonitored target was perceived to be more honest than the control target ($M = 0.42$), $t(32) = 6.73, p < .001$. Finally, as in Study 1, ratings of the monitored target and the control target did not differ significantly, $t(32) = 0.17$.

The pattern of results among participants in the low incentive condition was very different. When the incentive to cheat was low, participants' ratings of the honesty of the unmonitored target ($M = 1.59$) and the monitored target ($M = 1.24$) did not differ significantly, $t(33) = 0.35$. In addition, both the unmonitored target and the monitored target were perceived as more honest than the control target ($M = 0.18$), $t(33) = 5.64, p < .001$, and $t(33) = 3.92, p < .001$, respectively.



Figure 1. Participants' ratings of the monitored, unmonitored, and control targets along the dimensions of honesty, dishonesty, and ability.

Evidence of the Counterfactual Correspondence Bias

To determine whether the monitored target was perceived to be more dishonest than the control target, we followed the same procedure as in Study 1. Specifically, we examined participants' choice for future supervision and their nominations for the most dishonest target.

Choice for supervision. To assess the perceived dishonesty of the monitored and control targets, we first compared the number of nominations each target received when participants were asked to indicate which of the two targets they would choose to supervise in the future. Choice of future supervision varied as a function of incentive to cheat, $\chi^2(1, N = 67) = 7.89, p < .01$. As expected and consistent with the results of Study 1, participants who observed the monitored target behave honestly when the incentive to cheat was high tended to prefer to supervise the monitored target rather than the control target in a future testing situation (64% vs. 36%), though this difference did not quite reach statistical significance, $\chi^2(1, N = 33) = 2.46, p = .12$. Participants who observed the monitored target behave honestly when the incentive to cheat was low, on the other hand, preferred to supervise the control target in the future (29% vs. 71%), $\chi^2(1, N = 34) = 5.77, p < .02$.

The finding that participants preferred to supervise the control target rather than the monitored target in the low incentive condi-

tion was unexpected. We had expected that participants would show no strong preference for supervising one target over the other in this condition. We return to this finding later.

Predictions about dishonesty. We next compared the dishonesty index (Cronbach's $\alpha = .85$) for the three targets by means of a repeated measures ANOVA, with surveillance as a within-subject factor and incentive level as a between-subjects factor. As predicted, this analysis revealed a significant interaction between surveillance status and level of incentive on the target's dishonesty index, $F(2, 64) = 3.89, p < .03$ (see Figure 1).

Replicating the results of Study 1, when the incentive to cheat was high, the monitored target ($M = 2.06$) was assigned a higher dishonesty score than the unmonitored target ($M = 0.33$), $t(32) = 5.54, p < .001$. In addition, as in Study 1, the monitored target was also perceived to be more dishonest than the control target ($M = 0.61$), $t(32) = 3.99, p < .001$. Ratings of the unmonitored target and the control target did not differ, $t(32) = 1.20$.

Once again, however, the pattern was very different among participants in the low incentive condition. When the incentive to cheat was low, the monitored target ($M = 1.18$) was perceived as no more dishonest than the unmonitored target ($M = 0.62$) or the control target ($M = 1.21$), $t(33) = 1.52, ns$, and $t(33) = 0.07, ns$, respectively.

Ability Attribution

Level of incentive also interacted with surveillance status in determining participants' estimates of the number of math items the targets answered correctly, $F(2, 64) = 9.29, p < .001$ (see Figure 1).

The results for the high incentive condition closely parallel those of Study 1 (see Figure 1). First, participants estimated that the unmonitored target's score ($M = 8.64$) would be higher than the monitored target's score ($M = 6.39$), $t(32) = 4.13, p < .001$. In addition, participants tended to predict that the unmonitored target had scored higher than the test average, $t(32) = 1.75, p < .10$, and that the monitored target had scored significantly lower than both the test average and the control target ($M = 7.85$), $t(32) = 3.81, p = .001$, and $t(32) = 3.60, p = .001$, respectively.³

The results were strikingly different in the low incentive condition. The monitored target in this condition was predicted to perform significantly better ($M = 9.29$) than both the test average and the control target ($M = 7.97$), $t(33) = 2.98, p < .01$, and $t(33) = 2.68, p = .01$, respectively, and comparable to the unmonitored target ($M = 8.88$), $t(33) = 0.41, ns$.

Counterfactual Thought and Target Evaluation

We again examined the correlation between participants' estimates of the likelihood that the monitored target would have cheated if he had not known he was being videotaped and their evaluation of him. Consistent with our predictions, the more likely participants thought it was that the monitored target would have cheated if not monitored, the higher they rated him on the dishonesty index ($r = .50, p < .001$) and the lower they tended to perceive his math ability to be ($r = -.21, p < .10$). Moreover, as in Study 1, participants' evaluations of the unmonitored target were also related to their theories about the situation: The more likely participants thought it was that the monitored target would have cheated if he had not been monitored, the higher they rated the unmonitored target on the honesty index ($r = .43, p < .001$) and the higher they perceived his math ability ($r = .27, p < .03$).⁴

Finally, we explored the plausibility of the psychological process that we believe is responsible for the observed pattern of results, using procedures outlined by Baron and Kenny (1986) for assessing mediation. Specifically, we tested the notion that the high incentive to cheat increased the likelihood that participants would endorse the belief that the monitored target would have cheated if not for the supervision, which, in turn, increased the perception that the monitored target was dishonest. As reported above, the incentive manipulation significantly predicted participants' endorsement of the counterfactual ($B = -1.43, SE = 0.54, p = .01$). In addition, as we also reported above, endorsement of the counterfactual was a significant predictor of participants' ratings of the monitored target on the dishonesty index ($B = -0.48, SE = 0.13, p < .001$). Finally, when both incentive and endorsement of the counterfactual were included in the same equation predicting perceived dishonesty, the former dropped to nonsignificance, and the latter remained highly significant ($B = 0.92, SE = 0.61$, and $B = -0.50, SE = 0.13, p < .001$, respectively). Results of a Sobel (1982) test confirmed the significance of this mediated relation ($z = 2.07, p < .04$).

Discussion

Study 2 went beyond Study 1 by manipulating the incentive to cheat along with the constraints against doing so. The manipulation of incentive had the predicted effect on participants' counterfactual thoughts and their judgments of the target's honesty. First, participants indicated that there was a greater likelihood when the incentive to cheat was high than when it was low that the monitored target would have cheated if he had not been monitored. Second, despite identical behavior, participants judged the monitored target as more dishonest when the situation in which he was monitored presented him with a high as opposed to a low incentive to cheat.

The one unpredicted finding to emerge from Study 2 pertains to the monitored target in the low incentive condition. Not only was this target viewed more positively than his high incentive counterpart (as expected), but he was viewed more positively than the control target in the low incentive condition. Although not predicted, this finding may also reflect the tendency of observers watching a monitored target to imagine what the target would have done if he or she did not know he or she was being monitored. The difference in the low incentive condition, however, is that in this condition observers assumed that the target also was highly unlikely to have cheated under unmonitored conditions ($M = 3.63$ on a 9-point scale), a belief that might have led observers to conclude that the target was more honest than average. In other words, observers in both the high and the low incentive conditions acted as though they had two pieces of information about the monitored target: what the target did do (not cheat in both situations) and what he likely would have done if he did not know he was being monitored (cheat in the high incentive situation and not cheat in the low incentive situation). The counterfactual information that participants had in the high incentive situation had the consequence of making the target seem more dishonest than average, whereas in the low incentive situation it had the consequence of making him seem more honest than average.

Study 3

We elicited participants' counterfactual thoughts about the target directly in Studies 1 and 2. That is, the dependent measures specifically instructed participants to consider counterfactual alternatives to the behavior that they had observed. We did this because our focus in these studies was the consequences of counterfactual thoughts (however prompted) on the person perception process. According to our analysis, however, situations such as the

³ Once again, the assumption that participants viewed the control target as a proxy for the average 10-year-old boy is supported by the closeness of participants' estimates of the control target's performance to the average test score, $t(32) = 0.54, ns$.

⁴ The correlations reported here are from the full sample, collapsed across conditions. Although there was a substantial mean difference in the degree to which participants in the two conditions endorsed the counterfactual, the logic of these associations applies equally to both conditions: Within each condition, participants who were relatively more confident that the monitored target would have cheated had he not known about the video camera should have formulated a more negative impression of him, and they should have been more impressed by the honest behavior of the unmonitored target. Consistent with this reasoning, the within-cell correlations were very similar to those for the full sample reported here.

high incentive conditions of Studies 1 and 2 naturally and spontaneously evoke counterfactual thoughts of cheating behavior in observers. Study 3 seeks support for this claim.

Method

Participants

Sixty members of the Princeton University community (including undergraduates, graduate students, and staff members) were paid \$8 for participating in this study.

Procedure

The experimental design of Study 3 is similar to that of Study 2. In the current study, however, all participants observed just one video clip of a 10-year-old boy taking a math test. For all participants, the situation presented strong constraints against cheating (i.e., the boy was aware of the video camera). In all other ways, the instructions were the same as those used in Study 2. For half of the participants, the situational incentive to cheat was described as high (i.e., insufficient time to complete the math problems, financial reward for each correct answer), and for the other half of the participants, the incentive to cheat was described as low (i.e., sufficient time to complete the problems, no financial reward for correct answers).

All participants watched the same video clip, following which they answered four questions that appeared on separate pages. The first question asked participants to list the thoughts they had as they watched the video. Specifically, they were asked to

please take a few minutes to write down all the thoughts that went through your mind as you watched the video clip. So do not worry about using proper grammar or spelling—just try to capture all of the thoughts that occurred to you while you were watching the video.

Participants were provided ample space to record their thoughts.

The second question requested participants to complete the sentence, “If this boy had not known about the hidden camera . . .” The third question was designed to assess whether witnessing the nondiagnostic behavior of the target led participants to conclude that he was more prone to dishonesty than the average child of his age. We accomplished this by informing participants that previous research had found that approximately 50% of boys of this age cheated if they were unaware of the hidden camera and by then asking them to indicate how likely it was that the boy they had observed would have been among the 50% who cheated if he had been unaware of the video camera. We described the odds as 50/50 to ensure that participants had no a priori basis for assuming that a child about whom they had no diagnostic information was or was not prone to cheating. Participants responded to this question using a 9-point scale with endpoints labeled *very likely* and *very unlikely* and the midpoint labeled *neither likely nor unlikely*.

This latter question substituted for the nomination questions used in Studies 1 and 2. The reason for this substitution is that because we did not include a control target, someone about whom the participants had no information, we needed an alternative means of determining whether participants had drawn an inference about the monitored target’s personal disposition toward dishonesty. No longer able to make this assessment by comparing participants’ perceptions of the monitored target with those provided for the control target, we needed a measure that by itself would indicate whether participants believed they had learned anything about the target’s personal disposition toward dishonesty. The new measure was assumed to accomplish this through its midpoint. Specifically, participants who felt that they had learned nothing about the dishonesty of a target (and thus could not say whether he was one of the 50% who cheated) were able to communicate this by checking the midpoint of the scale (i.e., indicating that they believed it was neither likely nor unlikely that he was one of the 50% who cheated).

The fourth and final question began by informing participants that “while watching this clip, some people report thinking to themselves, ‘If he did not know about the hidden camera, he probably would have cheated.’” Participants were then asked whether this thought had occurred to them as they watched the video.

After participants completed the questionnaire, they were probed for suspicion and debriefed.

Results

The first two dependent measures were open ended and did not ask directly about counterfactual thoughts pertaining to cheating. The second two questions, in contrast, asked directly about the likelihood that the target would have cheated if he did not know that he was being videotaped.

Open-Ended Questions

Two raters who were blind to condition coded the responses to the two open-ended questions. The first of these questions (i.e., “Write down all the thoughts that went through your mind as you watched the video clip”) was designed to assess the extent to which participants spontaneously generated counterfactual thoughts of cheating behavior. To this end, the coders were instructed to identify all statements that made reference to the target being tempted to cheat or being inhibited from cheating because of the camera. Examples of responses placed in this category are “Maybe if he did not know the camera was there he would cheat” (Participant 13) and “I also wondered if he would have looked at the answers had he not known that there was a camera taping him” (Participant 7).

An examination of the thought listings revealed two other common categories of responses. The first of these involved suggestions that the behavior of the target was being influenced by the presence of the video camera (e.g., “He’s definitely aware of the camera and seems to be playing to it”; Participant 9). The second involved references to the target appearing nervous or anxious (e.g., “During this time he became fidgety and looked around the room and up occasionally”; Participant 24). Raters coded for these categories as well. The interrater agreement was very high across all of the statements, ranging from 86% to 93%. The few discrepancies that emerged were resolved through discussion.

Consistent with the hypothesis, participants were more likely to spontaneously generate counterfactual cheating thoughts in the high than in the low incentive condition. Specifically, 63% of the participants in the high incentive condition made one or more such statements, whereas only 23% did so in the low incentive condition, $\chi^2(1; N = 60) = 9.77, p = .002$. There were no significant condition differences in the responses placed in the other two categories.

Coders also examined participants’ responses to the second open-ended question, which directly invited participants to generate counterfactual thoughts about how the target would have behaved if he had not known about the video camera. Responses were coded 0 if they made no mention of cheating behavior, 1 if they expressed the thought that the boy might have cheated if he had not known about the camera, and 2 if they expressed the definite belief that the boy would have cheated if he had not known about the camera. Given that this question more explicitly focuses on counterfactual thoughts, it is not surprising that more of the responses made reference to the target cheating. Nevertheless, despite the elevated number of references to cheating behavior generally, the predicted condition difference still

emerged. Eighty-seven percent of the participants in the high incentive condition completed the sentence stem with reference to cheating (47% said he definitely would have cheated, and 40% said he might have cheated). In contrast, 43% of participants in the low incentive condition said that the boy might have cheated if he had not known about the hidden camera, and none of these participants said that he definitely would have cheated, $\chi^2(1, N = 60) = 22.08, p < .001$.

Closed-Ended Questions

With two closed-ended questions, we directly tapped participants' thoughts about the likelihood that the target would have cheated if he did not know of the camera. The first of these asked participants how likely (on a 9-point scale, with higher numbers indicating higher likelihood) they thought it was that this boy would be one of the 50% of boys who typically cheat when they do not know they are being observed. As predicted, participants in the high incentive condition thought it was more likely the target would be one of the 50% of boys who cheated than did participants in the low incentive condition ($M_s = 6.50$ vs. 4.50), $t(58) = 5.81, p < .001$. Furthermore, the judgments made by participants in the high incentive condition were significantly higher than the neutral midpoint (labeled *neither likely nor unlikely*), whereas the judgments made by participants in the low incentive condition were significantly lower than the midpoint, $t(29) = 6.30, p < .001$, and $t(29) = 2.11, p = .04$, respectively.

A closer examination of the responses to this question reveals just how differently participants saw the target in the two conditions. In the high incentive condition, 83% of the participants used a scale point that was above the midpoint, indicating that they thought it was more likely than not that the target would be included among the cheaters. Ten percent of participants in the high incentive condition selected the scale midpoint, and only 7% selected a scale point below the midpoint (which reflected the belief that the target was unlikely to be among the 50% who cheated). In short, participants in this condition thought they were observing a cheater. The pattern was very different in the low incentive condition. In this condition, only 17% of the participants chose a scale point above the midpoint, and more than twice that many (40%) chose a scale point below the midpoint (indicating that they thought it was unlikely that the target would have been among the 50% who cheated); 43% selected the scale midpoint. In short, participants in this condition saw no reason to assume that they were observing a cheater. A chi-square test revealed the distribution of responses to be significantly different in the two conditions, $\chi^2(2, N = 60) = 26.73, p < .001$.

The final question addressed the cheating counterfactual most explicitly. Participants were told that some people who viewed the video clip they had just seen reported thinking to themselves, "If he did not know about the hidden camera, he probably would have cheated," and they were asked whether this thought had occurred to them. The results in this case, too, were as predicted. Ninety percent of the participants in the high incentive condition said "yes" (10% said "no"), whereas only 3% of the participants in the low incentive condition said "yes" (97% said "no"), $\chi^2(1, N = 60) = 45.27, p < .001$.

Counterfactual Thought and Target Evaluation

Participants in the current study were explicitly told that the odds of any particular boy cheating on the test were 50/50, and they were

overtly invited to indicate (via the *neither likely nor unlikely* response option) that they did not have sufficient individuating information to judge whether the boy they had observed was inherently more or less likely than average to cheat if the opportunity presented itself. Yet, as we reported above, the vast majority of participants in the high incentive condition apparently felt that they did have sufficient information for making such a judgment: Over 80% of these participants indicated that it was more likely than not that this boy was a cheater. This dispositional judgment is similar to the trait inferences made by participants in Studies 1 and 2. In our final analysis, we explored the notion that the impact of the incentive manipulation on this trait inference was mediated by the counterfactual thoughts about the target that participants spontaneously generated as they watched the video clip.

Consistent with the results reported above, a dummy variable contrasting the low and high incentive conditions significantly predicted the negative dispositional inference about the target ($B = 2.00, SE = 0.34, p < .001$). In addition, as we reported above, the incentive manipulation was also a strong predictor of the spontaneous generation of the thought that the target would have cheated if he had not been aware of the camera ($B = 0.40, SE = 0.12, p = .001$). When both incentive and spontaneous generation of the counterfactual were included in the same equation, both remained significant predictors, though the magnitude of the former predictor dropped slightly ($B = 1.71, SE = 0.37, p < .001$, and $B = 0.72, SE = 0.37, p < .06$, respectively). Results from a Sobel (1982) test confirmed that this mediated relation was marginally significant ($z = 1.68, p = .09$). If, in place of participants' purely open-ended responses, we use their counterfactual stem completions (i.e., "If this boy had not known about the camera . . .") as a potential mediator, the magnitude of the relation between incentive level and negative trait inference drops more substantially, and stem completions remain a significant predictor of trait inferences ($B = 1.46, SE = 0.42, p = .001$, and $B = 0.60, SE = 0.28, p = .03$). A Sobel test confirmed that this mediated relation was significant ($z = 2.00, p < .05$).

Discussion

We hypothesized that the reason participants in the high incentive conditions in Studies 1 and 2 inferred that the noncheating monitored target was disposed to cheat was that they found themselves imagining the child cheating under different circumstances as they watched the tape. The present results strongly support this claim. On four different measures, participants viewing the monitored target in the high incentive condition revealed a greater incidence of counterfactual thoughts pertaining to cheating than did participants in the low incentive condition. This difference emerged both on measures that directly elicited counterfactual thoughts and on nondirective, open-ended measures. Most impressive, 63% of participants in the high incentive condition spontaneously made reference to counterfactual thoughts of cheating behavior when asked simply to indicate the kinds of thoughts the video clip evoked.

In addition to providing evidence that observers who watched a child who did not cheat spontaneously imagined the child in another situation in which he would cheat, Study 3 also replicates, using a different measure, the principal findings of Studies 1 and 2. Participants in the high incentive condition concluded not only that the boy they watched was more likely to cheat than partici-

pants in the low incentive condition thought but that he was more likely to cheat than was the average boy. More specifically, despite the fact that participants were presented with base rate information suggesting that the odds of a boy this age cheating under these circumstances were 50/50 and despite the fact that participants were explicitly offered a response option expressing the view that this particular boy was neither likely nor unlikely to be a cheater, participants in the high incentive condition overwhelmingly speculated that the boy they observed was likely to be one of the 50% who cheat when they do not think they are being monitored. In contrast and also replicating Study 2, participants in the low incentive condition were more inclined to speculate that the boy they observed was likely to be among the 50% who refrain from cheating even when they are unaware that they are being monitored. In short, participants concluded that a child they observed in a situation that prompted counterfactual thoughts of dishonesty was a dishonest child and that a child they observed in a situation that prompted no such thoughts was an honest child.

General Discussion

The present research focused on the relations among surveillance, temptation, and perceived honesty. The target event in Studies 1–3 involved a 10-year-old boy who resisted the temptation to cheat when he knew that he was under surveillance. Consistent with Strickland's (1958) classic finding, observers rated this target as less honest than a target who behaved similarly but who did not know he was being monitored. In addition and most significant from the present perspective, observers rated the focal target as more dishonest than a control target, someone about whom they had no information other than a photograph and some meager demographic details.

Why might people assume that a child who behaved as they expected him to (i.e., resisted the temptation to cheat) is more dishonest than average? The explanation, we suggest, lies with the counterfactual norm observers generated while observing the target. To begin with, observers believed that the average 10-year-old boy would be very tempted to cheat in the high incentive condition of Studies 1–3 if he did not think he would get caught. This expectancy, quite reasonably, led observers watching an unmonitored target resist a strong temptation to cheat to conclude that he was more honest than average. On the other hand, and less reasonably, it also led observers who watched a monitored target resist the same temptation to generate the counterfactual or post-computed expectancy that he would have cheated if he did not think he would get caught (Study 3). Although observers did not actually see the target turn around and look at the answers on the board, they saw him do this in their mind's eye. As a consequence of the counterfactual they generated, observers had "knowledge" about the target that went beyond the knowledge that he did not cheat: They knew that he surely would have cheated were he not being monitored. The situation was very different when the incentive to cheat was low. When observers in this situation were prompted to contemplate what the target would have done under unmonitored conditions, they most likely imagined him not cheating there as well. This "knowledge" led them to conclude that this target was more honest than the average 10-year-old.

Relation to Other Models of Person Perception

The present analysis's emphasis on the perceiver's theory of the situation is shared by virtually all person perception models. The roles that perceivers' expectations play in other models are quite different from the role assumed here, however, and these differences warrant closer examination.

Consider first Kelley's (1971) influential analysis of the role that causal discounting plays in the person perception process. From Kelley's perspective, perceivers' theories of the situation are important because the person information the perceivers extract from the actions of others depends on how they parse causality for that behavior between the disposition of the actor and the features of the situation. The more consistent the target's behavior is with what perceivers judge to be the press of the situation, the more they will discount the role played by factors internal to the target. In the present context, this means that the more consistency participants saw between the target's behavior and their theory of what 10-year-old boys will do in that situation, the more they should have discounted the causal role played by the target's internal dispositions in that behavior. Indeed, the finding that participants attributed greater honesty to a noncheating unmonitored target than to his monitored counterpart suggests that participants did discount the causal role played by dispositional honesty more in the monitored than in the unmonitored condition.

However, what about the finding that perceivers attributed greater dishonesty to the monitored target than to the control target? Can this result be accommodated within Kelley's (1971) discounting analysis? It certainly does not seem to represent a rational application of discounting logic. However, neither does it represent an instance of the well-documented phenomenon of underdiscounting, wherein perceivers attribute a target's behavior (e.g., noncheating) to a corresponding disposition (e.g., honesty) even when the behavior is consistent with situational constraints (Gilbert, 1998). For their attributions to qualify as a case of underdiscounting, perceivers would have had to judge the monitored (noncheating) target to be more honest, not more dishonest, than the control target.

The question then arises as to whether the greater dishonesty attributed to the monitored than to the control target could be described as reflecting a case of overdiscounting. The answer to this question, too, appears to be no. For even if perceivers did overdiscount, thereby underestimating the causal role that the target's dispositional honesty played in his honest behavior, this would not explain the greater dispositional dishonesty attributed to the monitored target. Overdiscounting perceivers could be expected to draw insufficiently strong inferences of honesty from the monitored target's behavior but not overly strong inferences of dishonesty. Stated differently, the striking finding is not that perceivers drew too little person information from the monitored target's behavior, as would be the case if they had fallen prey to overdiscounting; it is that they drew too much.

In another influential analysis, Trope (1986) proposed that there are two stages in the person perception process and that perceivers' theories of the situation play a different role in each. With respect to the first stage, or what Trope termed the identification stage, perceivers use their theories along with other prior beliefs to spontaneously code the observed behavior in terms of disposition-relevant categories. At the second, or inference, stage, perceivers use their theories to adjust from the disposition implied by the identified behavior to diagnose the degree of dispositional information in that behavior. The

correspondence bias emerges, according to Trope, from the fact that the subtractive second stage often does not sufficiently compensate for the additive impact that perceivers' theories have on the coding of the behavior at the first stage.

In light of Trope's (1986) analysis, a reasonable question to ask is whether the dishonesty attributed to the monitored target might reflect the impact of perceivers' theories on their encoding or construal of the target's behavior rather than on the content of the counterfactual thoughts the target's behavior brought to mind. That is, might the tendency of participants to see the monitored target as dishonest reflect their theory-biased coding of the data they "saw" on the tape, not the additional (counterfactual) data that their theories helped generate in their imagination? There are a number of conceptual and empirical reasons to doubt that this is the case. First, the instruction to all participants that they would not actually see incidents of cheating on the tapes should have discouraged rather than encouraged the coding of behavior in dishonesty-consistent terms. Moreover, if participants coded the actions of the noncheating targets in dishonesty-consistent terms, this seems more likely to have taken place when they were watching the unmonitored target than when they were watching the monitored target. The reason for this is that the expectation that the boy they were watching would cheat sometime during the session (though not on the taped period that they were viewing) would seem more reasonable in the unmonitored than in the monitored condition.

Third, if participants were disposed to see more signs of dishonesty in the behavior of the monitored than the unmonitored target, it seems logical that this would occur, albeit to a lesser extent, in the low incentive as well as the high incentive condition. The fact that the monitored target in the low incentive condition was seen as more honest (not dishonest) than the control target is inconsistent with this possibility, though, as noted earlier, it is consistent with the claim that participants in this condition were influenced by the counterfactual thought that the target would not have cheated even if he were unaware that he was being monitored.

Finally, analyses of the content of the thought-listing data (Study 3) revealed no evidence that participants coded the behavior they saw on the tape differently in the monitored and unmonitored conditions. First, there was no difference in the number of thoughts listed in the two conditions. Second, relatively few participants in either condition (less than 12%) generated thoughts that suggested they had coded what they saw in dishonesty-relevant terms. Third, there were no condition differences in the two most common response categories other than counterfactuals (nervousness and camera-sensitive behavior).

Why Did the Cheating Counterfactual Come to Mind?

The number of potential counterfactual thoughts that any particular behavior could evoke is considerable, to say the least. So why would the counterfactual evoked by the key social event featured in the present research (i.e., a boy resisting the temptation to cheat under high temptation and high likelihood of detection) involve the modification of one aspect of the situation (i.e., the likelihood of being detected) rather than another (e.g., the incentive to cheat)? One possibility is that observers generated the counterfactual that best promoted their goal in the situation. As previously documented, one function of counterfactuals is that they facilitate goal achievement (Roese & Olson, 1995). If one

assumes that observers in the present context had the goal of assessing the honesty of the targets, then they might reasonably (as the opening quote from Mead, 1934, describes) try to imagine the target in a situation that would more effectively diagnose honesty. Certainly, if observers could actually request additional information about the target, asking for information about the target's behavior in a more diagnostic situation would seem eminently reasonable. In short, whether one is speaking of counterfactual data or actual data, the data that are most diagnostic of the monitored target's honesty seem to be generated by the situation in which the target was confronted with the same level of temptation (high or low) but low (rather than high) likelihood of detection.

However, why were the observers of the monitored target in the present experiments motivated to assess the target's honesty? One possibility is that the experimental context induced this goal in participants. Another possibility is that the goal of honesty detection, or, more accurately, dishonesty detection, is a chronically activated one for observers. This latter possibility is supported by evidence suggesting that people tend to be highly vigilant for signs of dishonesty (Fein, Hilton, & Miller, 1990; Hilton, Fein, & Miller, 1993). Of the various errors that observers can commit, being duped may be the one they fear most. For example, when observers have even the slightest ground for questioning the authenticity of the actor's behavior, their vulnerability to the correspondence bias is greatly reduced or eliminated (Hilton et al., 1993). Concluding that an introvert is actually an extrovert (or that an honest person is actually dishonest) makes one wrong, but concluding that a dishonest person is actually honest makes one a fool. Another piece of evidence for people's special sensitivity to signs of dishonesty is the asymmetrical weight they accord evidence of honesty and dishonesty in the impressions they form of others. Committing a single act (and possibly even a single counterfactual act) of dishonesty is sufficient for an actor to find himself or herself being labeled a dishonest person (Reeder & Brewer, 1979). In contrast, a single act (or counterfactual act) of honesty is insufficient to establish a reputation for honesty. Humans' hypersensitivity to evidence (including possibly counterfactual evidence) of dishonesty may even, according to some, have been favored by evolutionary pressures (Cosmides, 1989; Shackelford & Buss, 1996). Detecting violations of social norms is so important to social life, the argument goes, that humans have evolved a "cheater detection mechanism" (Cosmides, 1989, p. 188).

Positing that observers had the goal of assessing the target's honesty may be neither sufficient nor necessary to explain the present results, however. Imagine that the present experiment were described as being about the detection of extroversion rather than cheating behavior. Would we expect people who watched a person act sedately in an introversion-inducing situation to generate thoughts of the person in a more extroversion-inducing situation? Possibly, but it does not seem likely. For one thing, it is difficult to think of a situation in which, although introversion is the norm, it is easy to mentally modify the situation into one in which extroversion is the norm. In the present experimental context, by contrast, not only does it seem easy to imagine the monitored target in an unmonitored situation, it is difficult to suppress such thoughts. In summary, whether people have the goal of diagnosing the presence of a particular trait or not, unless the nature of the situation observed is easily imagined otherwise, they are unlikely to generate counterfactual images.

Consequences of Counterfactual Acts of Cheating

As interesting and important as is the question of how spontaneously people generate images of counterfactual cheating, it is not the major question addressed in the present research. Of primary concern here are the consequences of generating such images. Specifically, the present research asks whether observers who, for whatever reason, generate counterfactual thoughts of dishonesty about a person are disposed to see that person as dishonest. The answer to this question is clearly yes. Observers seem either not to recognize or to insufficiently adjust for the fact that their evidence that the monitored target is dishonest stems solely from their theory about what the average person would do. In either case, they are guilty of conflating evidence and theory.

Implications

That observers are influenced by counterfactual acts of deceit may be one reason people so dislike being supervised or chaperoned. Targets dislike being supervised no doubt in part because of what it implies about the supervisor's presupervision perception of the target—he or she needs supervision. However, they may also dislike it because of its assumed effect on the supervisor's postsupervision perception of the target—he or she, often despite exemplary behavior, is seen as even more in need of supervision than was originally assumed. The chaperone's growing distrust of his or her charges is not likely to be lost on the charges. Furthermore, the source of this frustration is not simply the recognition that being chaperoned will make it difficult to disconfirm the chaperone's suspicions. It is the recognition that being chaperoned will lead to the confirmation—indeed, the strengthening—of the chaperone's suspicions.

From the supervisor's or even the impartial observer's perspective, the experience of witnessing someone behave lawfully or honestly under close supervision is likely to have implications for beliefs that go well beyond the particular party or parties observed. The "witnessing" of counterfactual sinning in one person is likely to make observers more cynical about people in general. Thus, people who supervise others in tempting situations might be expected to acquire highly cynical views about the general level of honesty in the population, even in the absence of any actual evidence of a single person's dishonesty. Indeed, people's cynicism may be more a function of how long they have supervised or monitored people (and hence how much counterfactual evidence of dishonesty they have) than of the amount of dishonesty they have witnessed. In this respect, people's beliefs about general inclinations toward dishonesty become self-reinforcing. That is, these views, although themselves generated by observers' general beliefs about people, come, through the counterfactual thoughts they generate about a particular target, to strengthen observers' beliefs about people in general. Thus does cynicism beget suspicion, which, in turn, begets greater cynicism. Economists describe situations in which there are no constraints against pursuing collectively costly self-interest (e.g., unmonitored highway driving speed) as creating a moral hazard. The present research suggests that there is also hazard created when one does monitor people in tempting situations: the hazard that an unnecessary policy of constraints will be perpetuated. Whether one is judging personal politics or public policies, it is one thing to condemn sinners; it is quite another to condemn counterfactual sinners.

References

- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
- 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.
- Fein, S., Hilton, J. L., & Miller, D. T. (1990). Suspicion of ulterior motivation and the correspondence bias. *Journal of Personality and Social Psychology, 58*, 753–764.
- Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 89–150). New York: McGraw-Hill.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Hilton, J. L., Fein, S., & Miller, D. T. (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin, 19*, 501–512.
- Jones, E. E. (1990). *Interpersonal perception*. New York: Freeman.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266.). New York: Academic Press.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.
- Kelley, H. H. (1971). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.
- Mead, G. H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.
- Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 37, pp. 233–256). Palo Alto, CA: Annual Reviews, Inc.
- Olson, J. M., Roesse, N. J., & Zanna, M. P. (1996). Expectancies. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 211–238). New York: Guilford Press.
- Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution interpersonal perception. *Psychological Review, 86*, 61–79.
- Roesse, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133–148.
- Roesse, N. J., & Olson, J. M. (1995). Functions of counterfactual thinking. In N. J. Roesse & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 169–197). Mahwah, NJ: Erlbaum.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 174–221). New York: Academic Press.
- Shackelford, T., & Buss, D. M. (1996). Betrayal in mateships, friendships and coalitions. *Personality and Social Psychology Bulletin, 22*, 1151–1164.
- Strickland, L. H. (1958). Surveillance and trust. *Journal of Personality, 26*, 200–215.
- Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equation models. In S. Leinhardt (Eds.), *Sociological methodology* (pp. 290–312). San Francisco: Jossey-Bass.
- Trope, Y. (1986). Identification and inference processes in dispositional attribution. *Psychological Review, 93*, 239–257.

Received August 30, 2004

Revision received March 29, 2005

Accepted March 30, 2005 ■