
Suspicion, Affective Response, and Educational Benefit as a Result of Deception in Psychology Research

Nicholas Epley
Chuck Huff
St. Olaf College

This research evaluated participants' reactions to deception in experiments by having them participate in a replication of a deception experiment. Half of the participants were made aware of this deception immediately, whereas the other half were not. Participants reported little negative impact from being deceived, but significant negative effects were reported on the basis of receiving negative feedback (a manipulation in the deception experiment). Furthermore, participants who were informed of the deception became more suspicious than uninformed participants, and this effect lasted for 3 months after the initial experience. Thus, deception may not be as costly to participants as commonly believed. From a cost-benefit standpoint, other issues (e.g., suspicion and negative stimuli in experiments) should be of greater concern.

Despite repeated ethical criticism of deception in psychology experiments (Baumrind, 1985; Oliansky, 1991) and strong concerns about deception's effects on public perceptions of psychology (Kelman, 1967), deception has remained an extremely popular methodological tool. Seemans (1969) reported that only 18% of all studies in the *Journal of Abnormal and Social Psychology* and the *Journal of Personality* in 1946 involved deception. This percentage increased markedly by 1963, with 38% being deceptive studies in these journals. Between 1971 and 1974, approximately 57% of the studies reported in the *Journal of Personality and Social Psychology (JPSP)* (McNamara & Woods, 1977) contained deception, and in 1979, 58.5% of the studies reported in that journal were deceptive (Adair, Dushenko, & Lindsay, 1985). Our own review of the 1996 volume of *JPSP* showed that 42% of the articles reported studies involving deception, a decrease from earlier years but still substantial.¹ Deception has clearly become, and remains, a standard tool in the field of psychology.

Research investigating opinions about deception consistently reveals that faculty members and researchers are more concerned about the dangerous effects of deception on participants than are the participants themselves (Collins, Kuhn, & King, 1979; Fisher & Fyrberg, 1994; Korn, 1987). In response to these concerns, several psychologists have attempted to answer empirically the question, "How do actual participants feel about being deceived in psychological research?" (Fleming, Bruno, Barry, & Fost, 1989; Gerdes, 1979; Smith, 1981; Smith & Richardson, 1983).

This empirical approach to the investigation of ethical issues is prompted by what philosophers call a consequentialist approach to ethical decision making (Pettit, 1991). This is a normative approach that equates what is right or moral with what brings about the most positive consequences. If we use the consequences of our actions to judge their moral worth, then information about those consequences is obviously relevant when determining what is moral and what is not. The American Psychological Association (APA) endorses this approach when it states that "the investigator has a special responsibility to determine whether the use of such techniques is justified by the study's prospective scientific, educa-

Authors' Note: Nicholas Epley is now at the Department of Psychology, Cornell University. We would like to thank Richard Moreland, Kenneth Savitsky, Mark Sundby, and three anonymous reviewers for their many helpful comments on previous drafts of this article. We would also like to thank Rolf Skar for his assistance running participants and administering questionnaires and Cynthia Herrick and Joanna Hirvela for their long hours calling participants and coding data. Correspondence concerning this article should be addressed to Nicholas Epley, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14850 or Chuck Huff, 1520 St. Olaf Avenue, Northfield, MN 55057-1098; e-mail: ne12@cornell.edu or huff@stolaf.edu.

PSPB, Vol. 24 No. 7, July 1998 759-768
© 1998 by the Society for Personality and Social Psychology, Inc.

tional, or applied value" (APA, 1982, pp. 35-36). Thus, the reactions of research participants to being deceived are of crucial importance in this analysis of consequences. Such reactions can be used to determine if the gain in insight for the scientific community is worth the cost (inconvenience, humiliation, anger, etc.) to participants.

An analysis of consequences, however, cannot answer the concerns of those who argue against deception from a deontological or rights-based perspective (Krupat & Garonzik, 1994). Deontological arguments (from the Greek *deon* meaning duty) against the use of deception are not based on consequences but rather on concerns for human rights. Arguments against deception based solely on the rights of participants or duties of researchers are therefore unaffected by any appeal to consequences, even if participants say they prefer and enjoy deceptive experiments. Most of the arguments in the literature, however, are not based on this position but are based on a consequentialist approach that is similar to that presented by the APA (e.g., Christensen, 1988; Fisher & Fyrberg, 1994; Gerdes, 1979).

This emphasis on the consequences of research has driven psychologists to argue that empirical evidence ought to be employed when making ethical cost-benefit analyses (Aitkenhead & Dordoy, 1985). Such investigations have yielded very interesting, and at times conflicting, results. Lab lore surrounding the use of deception suggests that participants in deceptive research enjoy their participation more than if it had not employed deception, possibly because research involving deception is simply more interesting and fun (Coulter, 1986). In addition, deceptive research often includes a more detailed debriefing, leading some psychologists to believe that participants in such research receive more educational benefit from this added information. Dissonance reduction (Festinger, 1957) may also be responsible for these effects because positive attitude change may occur if the situation is initially considered aversive. As Aronson and Carlsmith (1963) have shown, suffering can often lead to liking.

Whatever its cause, the claim that participants do not view deception negatively has been supported by data that reveals that participants typically report little negative affect after being deceived, that they enjoy their participation more than those who did not participate in deception experiments, and that they feel they have received more educational benefit from their experience (Fleming et al., 1989; Gerdes, 1979; Smith, 1981; Smith & Richardson, 1983).

Research on the Effects of Deception on Participants

Affective response to deception. On the surface, it seems counterintuitive to suggest that research participants may enjoy deception. However, in one of the seminal

studies in this area, Smith and Richardson (1983) administered a questionnaire to 464 students who participated in at least one psychology experiment at some point during a semester. Participants in deceptive research actually reported liking the research more than those who did not participate in any deceptive research. Furthermore, only 28% of the participants in deceptive research reported feeling harm at any time, and much of that harm was alleviated by effective debriefings.

Given the range of severity involved in different kinds of deceptive research, one may wonder whether results like these apply equally well to more severe types of deception, such as those involving false feedback or staged emergencies (Rubin, 1985). However, similar results have been documented across a wide range of deceptive methods, including false emergencies (Schwartz & Godlieb, 1981), withholding of relevant information (Fleming et al., 1989; Gerdes, 1979; Smith, 1981), and false performance feedback (Smith, 1981). It is interesting to note that only 1.3% of the participants in Milgram's (1964) obedience experiments reported any negative feelings about their experience, and much of this negative affect may have been due to the outcome of the experiments rather than to the actual deception involved (Berscheid, Baron, Dermer, & Libman, 1973).

Educational benefit. Smith and Richardson (1983) also found that participants who were deceived reported more educational benefit from their experience than did those who participated in nondeceptive research. One plausible cause for this effect is that debriefings in nondeceptive research are generally more terse and less interesting. In research involving deception, however, debriefings tend to be highly personal and detailed, containing more information and providing participants with more material from which to make judgments about educational benefit (Coulter, 1986). Thus, participants may not actually receive more educational benefits from deception research but simply claim such benefits because they remember that extensive information was presented to them (although they may not remember anything else about that information). There has not, however, been any research that independently measures the *actual* amount of educational benefit associated with participation in deceptive versus nondeceptive research.

Suspicion. In 1967, Herbert Kelman issued a warning to the field of psychology that deception may be harming psychological research by creating an inherently suspicious pool of participants. The continued use of deception, argued Kelman, creates an evermore sophisticated pool of participants who are keenly aware of the amount of deception that occurs in psychology research. As time passes, potential participants will become gradually

more and more distrustful of psychologists. After entering the laboratory, these wily participants will be highly suspicious of any manipulation, creating an even more artificial environment than already exists. It was Kelman's contention that suspicious participants would not behave normally in research settings and that the use of deception was potentially undermining the validity of all future psychological research.

Nearly 30 years after Kelman's (1967) warning, it remains unclear whether this prophecy has been fulfilled. Empirical research on the issue has provided mixed results. MacCoun and Kerr (1987) report an experimental incident in which someone acting as one member of a six-person mock jury lapsed into an actual grand mal seizure. When the other five participants were contacted a short time later and asked whether they believed the seizure was actually part of the experiment, three of them thought that it was. Interestingly, these three participants all had previous course work in psychology, as well as research participation experience, whereas the only participant who immediately attempted to help the victim had no previous course work in psychology. Although this is anecdotal evidence, it is a gripping description of the difficulties we might face if Kelman's (1967) prediction is correct.

Research investigating Kelman's (1967) hypothesis more directly, however, has yielded inconclusive results. Sharpe, Adair, and Roese (1992) investigated the extent to which participants were becoming more suspicious of psychological research, based on comparisons of suspicion in 1970, 1989, and 1990. Although participants in 1990 were slightly more suspicious than participants in 1989, and participants in 1989 were slightly more suspicious than participants in 1970, these differences were nonsignificant. However, Smith and Richardson (1983) found that those people who participated in deceptive research perceived psychologists as less trustworthy than did participants in nondeception research.

Need for Current Research

Research investigating the effects of deception on participants presents a muddled picture. First, it is unclear to what extent participants' self-reports of educational benefit are valid. Second, it is unclear whether suspicion is permanently increased among people who participate in deceptive research and what effects this might have on other research using the same pool of participants. Finally, research on the effects of deception versus no deception (Aitkenhead & Dordoy, 1985; Smith & Richardson, 1983) has actually investigated two groups of people participating in two completely different types of research projects. Thus, the results obtained from this work are confounded by possible extraneous variables associated with differences in research method-

ology, debriefing content, debriefing length, and so forth. In addition, much of this work employs retrospective self-reports that were conducted several months or even years after the initial experience (Pihl, Zacchia, & Zeichner, 1981; Smith, 1981; Smith & Richardson, 1983).

This research makes the participants' knowledge of the deception involved in an experiment the only variable that differs between groups of participants. We replicated the self-serving bias in causal attribution (Alloy & Abramson, 1979, 1982) by providing participants with false performance feedback and asking them to attribute their performance to a number of causal factors. At the end of the first session, only half of the participants were informed about the deceptive nature of the feedback. The remaining participants were informed at the end of the second session of the study, 3 to 5 days later. Reports of suspicion, affect, and educational benefit were collected at the end of the first session. In the second session, we measured participants' memory for the design and theory of the first session. This was intended to serve as an actual measure of educational benefit. All participants were then fully debriefed. Finally, participants' suspicion, affect, and perceived educational benefit were measured again 3 months after the final session.

METHOD

Overview

This experiment used a 2 (debriefing: partial, full) \times 2 (performance feedback: positive, negative) \times 2 (time of measurement: postexperiment, 3-month follow-up) factorial design, with the last factor varying within participants. The experiment was completed in three sessions spanning a 3-month time period. The first session involved completion of several tasks presented on a computer through a program created using Hypercard. One of the major purposes of this program was to replicate the self-serving bias in causal attribution. Using this program, participants completed a reading-comprehension task and were given false performance feedback (either positive or negative). Participants were then asked to complete an eight-item internal-external attribution questionnaire.

At the end of the first session, half of the participants were given a full debriefing of the false feedback manipulation, whereas the other half only received a partial debriefing that did not explain the deceptive nature of the manipulation. Thus, half of the participants knew that they were participating in a deception experiment, whereas the other half did not. Response to the experiment was then measured through a 10-item questionnaire that was completed immediately after the debriefing. During the second session, held 3 to 5 days later, participants completed an unanticipated recall ques-

tionnaire and everyone was given a full debriefing. Participants were contacted again 3 months later and asked to respond to the same questionnaire that was completed after the first session.

Participants

A total of 18 male and 39 female St. Olaf College students received course credit for their participation.² All participants were tested individually and were randomly assigned to one of four experimental conditions.

Procedure

First session. Participants were informed that the experiment entailed completing several tasks administered using a computer. Each participant was led to a small room that contained one computer and in which he or she was allowed to work through a short program. In this program, participants were first asked to complete the 30-item Self-Efficacy scale (Sherer et al., 1982) that contained an added question, "I am concerned about my academic ability." This final item allowed for the screening of participants who might be seriously harmed by the negative academic performance feedback that is provided later in the experiment. Two participants indicated strong agreement with this item and were immediately dismissed from the experiment after a full debriefing. The next element of the program involved a series of six undergraduate-level essays that were paired with questions that appeared to measure reading comprehension. Participants were given 7 minutes to complete this task. This brief time period was selected to ensure that none of the participants could complete the task, thus limiting the formation of performance expectancies. Answers to these questions were never actually recorded.

Positive or negative false feedback was then presented, apparently generated by the computer. Participants were presented with a message that stated, "You finished X questions in the time allotted. Y [depending on condition] of your answers were correct. Your score is in comparison to the Z other individuals who have done this task on this computer in the last week. Based on the number of correct answers relative to other participants in this experiment, you have scored in the [manipulation here] percentile." X was an accurate count of the number of questions that were completed in the allotted time. Y was $X - 2$ (for success) or $X - 5$ (for failure). All negative scores were converted to 1. Z was a count of the number of participants who had used that computer in the study, with 27 added to make the comparison group seem large enough to be valid. In the bracketed space, participants were informed that they scored in either the "upper 84th" or in the "lower 47th" percentile. A manipulation check asking them to report

this feedback information back to the computer was included shortly after this manipulation.

Finally, to test the self-serving bias in causal attribution, participants completed an 8-item internal-external attribution questionnaire adapted from Arkin and Maruyama (1979). Participants rated the extent to which each factor influenced their performance using a 7-point Likert scale ranging from 1 (*no influence*) to 7 (*tremendous influence*). The seven factors included three internal attributions (personal ability, individual preparation, individual effort) and four external attributions (difficulty of task, physical or mental condition during time of task, time constraints, and luck). For the entire scale, $\alpha = .539$. General attribution patterns were obtained by averaging the responses to the 4 external and 3 internal attributions and then computing the difference. This scale ranged from -6 to $+6$, with positive scores indicating general external attribution patterns and negative scores indicating general internal attribution patterns (Arkin & Maruyama, 1979).

After completion of the computer program, participants were given one of two debriefings. In one condition, participants were given a detailed process debriefing (see Aronson, Ellsworth, Carlsmith, & Gonzales, 1990) and informed of the deceptive feedback manipulations. We call this the full debriefing condition. In the other condition, participants were not informed of the deception. We call this the partial debriefing condition.

Because length and detail of debriefing may influence participants' reactions to experiments (Coulter, 1986), careful attention was paid to make the debriefings in both conditions roughly equal in length. This standardization decreases the likelihood that any differences in participants' reactions to the experiment were caused by confounding factors such as debriefing length and detail, rather than the specific debriefing content.

After the debriefing, participants signed up for the second session. They were also presented with a written debriefing and a paper copy of the material that was presented over the computer and told that they could read over this material if they wished. They were asked to cooperate with the experimenters by not revealing the details of the study to other students until a specified date.

Before leaving, participants were asked to complete a short 10-item questionnaire that was supposedly to be used by the St. Olaf Research Ethics committee. This questionnaire was designed to measure suspicion, affective reactions to the experiment, and beliefs about the educational benefit participants received from their experience. Participants were instructed to mark the extent to which they agreed or disagreed with each statement on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Two items were related to suspicion. One asked participants to rate their agreement with the statement, "As a participant in future research, I will not be suspicious of any information presented by the experimenter." The other asked participants to rate their agreement with the statement, "Psychologists are trustworthy and honest." These two items were not significantly correlated ($r = .28$), however, and thus were analyzed separately.

The one item related to educational benefit asked participants to rate their agreement with the statement, "I feel I have received significant educational benefit from this research experience." This question is patterned after a reasonable self-report measure of educational benefit (Smith & Richardson, 1983).

Seven items were averaged together to provide an index of participants' general affective response to their research experience ($\alpha = .71$). The seven items were (a) I have only experienced positive feelings during the course of this experiment; (b) At no time in the experiment did I feel overly anxious, embarrassed, or nervous; (c) I have enjoyed my experience as a subject in psychology research; (d) I am more interested in psychology research now that I have been an active participant; (e) In the future, I will be more likely to participate in a psychology experiment because of my experience in this research project; (f) As a result of this experience, I would be willing to recommend research participation to one of my close friends; and (g) I feel that I was treated in an ethical manner at all times during this experiment.

Second session. During this session, held 3 to 5 days after the initial meeting, participants were informed that their task was to complete a memory questionnaire that contained 4 free-recall and 16 multiple-choice questions. For each of the 4 free-recall questions, participants were asked to write as much information as they could possibly remember regarding four events from the first session: the introduction period, the computer program, the order of tasks presented in the experiment, and the debriefing session. The 4 free-recall questions were coded by two independent raters who counted the number of correct details listed for each question ($\alpha = .995$). The 16 multiple-choice questions ranged across all areas of the first session, including the purpose, method, and content of the experiment. Reliability was relatively low for these items ($\alpha = .420$), making any effects difficult to interpret. Participants took approximately 20 minutes to complete the entire memory questionnaire (both free-recall and multiple-choice questions).

To create reasonably equal levels of motivation for this detailed task, participants were informed that a monetary prize (\$25) would be awarded to the individual who reported the largest amount of correct information,

which was awarded several weeks after the experiment was completed. Afterward, all participants were given a final process debriefing (Aronson et al., 1990) that described all manipulations and elements of deception in the experiment.

Third session. The final session was conducted over the telephone approximately 3 months after the initial session. Participants were asked the same questions as in the original affective questionnaire (completed after the first session) but in reverse order. Different experimenters than those in Sessions 1 and 2 were employed in an effort to reduce demand characteristics and experimenter expectancy effects.

RESULTS

Participants generally reported positive reactions to the experiment across all measures and conditions, and these reactions were not affected by their debriefing condition. In addition, these positive reactions increased across time. However, participants who received negative false feedback generally reacted less positively than did those who received positive false feedback. Deception had an effect on just one measure of suspicion, the one associated with future research participation, and no differences were found on any of the educational benefit measures. Gender had no effect on any measures (all $ps > .1$), and thus, all results were collapsed across this factor.

Participant Loss and Suspicion of Manipulations

During the debriefing following the first session, researchers investigated participants' levels of suspicion in the full debriefing condition using several questions (e.g., "Is there anything that didn't seem quite right in this experiment?" "Do you think there was anything more to this experiment than meets the eye?"). The same procedure was used following the second session for participants in the partial debriefing condition. No participants were suspicious about the actual nature of the study, and most were surprised when informed of its purpose.³

Two participants were dismissed early from the experiment because of their low scores on the Self-Efficacy scale (see the Methods section), which made it seem unwise to give them negative performance feedback. Two other participants were only able to complete the first session, and three failed to complete all of the dependent measures. These five participants were thus dropped from all analyses, leaving a total of 50 participants in all three sessions.

TABLE 1: Affective Response, Suspicion, and Educational Benefit

	Partial Debriefing				Full Debriefing			
	Positive Feedback		Negative Feedback		Positive Feedback		Negative Feedback	
	Post-Experiment (n = 13)	3-Month Follow-Up (n = 13)	Post-Experiment (n = 13)	3-Month Follow-Up (n = 13)	Post-Experiment (n = 12)	3-Month Follow-Up (n = 12)	Post-Experiment (n = 13)	3-Month Follow-Up (n = 13)
Affective response	5.58 (1.12)	5.96 (0.96)	4.46 (1.31)	5.25 (1.39)	5.35 (0.86)	6.12 (0.74)	4.61 (1.65)	5.42 (1.04)
Suspicion								
Suspicious of experiments	5.46 (1.45)	3.83 (1.47)	5.08 (1.50)	4.58 (2.11)	3.69 (1.32)	3.38 (1.76)	3.50 (1.95)	4.07 (1.82)
Suspicious of psychologists	5.73 (0.90)	5.82 (0.87)	5.25 (0.86)	5.7 (1.42)	4.85 (1.63)	4.38 (2.06)	5.54 (0.97)	5.23 (1.36)
Educational benefit								
Received educational benefit	5.46 (0.88)	5.33 (1.30)	5.15 (0.90)	5.92 (1.16)	5.31 (0.86)	4.85 (1.46)	5.00 (0.58)	5.50 (1.09)
Free-recall questions	24.18 (6.18)		24.58 (11.07)		17.07 (7.23)		27.32 (12.26)	
Multiple-choice questions	13.81 (2.18)		13.08 (3.06)		13.69 (1.79)		14.07 (2.47)	

NOTE: Standard deviations for each mean are reported in parentheses.

Manipulation Check

All participants correctly reported their performance feedback (47 or 84), indicating (at minimum) a clear awareness of the feedback manipulation.

Replicating the Self-Serving Bias

The self-serving bias was replicated. Participants who were presented with positive feedback displayed more internally oriented attributions ($M = -.185$, $SD = 1.07$), whereas participants presented with negative feedback displayed more externally oriented attributions ($M = .529$, $SD = 1.10$), $F(1, 49) = 5.408$, $p = .024$. Although neither of these means was significantly different from 0, they were significantly different from each other.

Affective Response to the Experiment

Unless otherwise noted, all subsequent analyses employed a 2 (debriefing: partial vs. full) \times 2 (performance feedback: positive vs. negative) \times 2 (time of measurement: postexperiment vs. 3-month follow-up) mixed model ANOVA. Means and standard deviations for participants' scores on the affective response index are reported in Table 1. It should be noted that participants generally gave positive responses to all of the items in that index, and thus, lower scores do not indicate negative affect but rather less positive affect.

Participants who received negative performance feedback reported less-positive affect overall ($M = 5.42$, $SD = 0.72$) than did participants who received positive feedback ($M = 5.86$, $SD = 0.64$), $F(1, 47) = 4.80$, $p < .05$. Participants also reported more positive affect after 3 months ($M = 5.80$, $SD = 1.11$) than they did after the first

session ($M = 5.41$, $SD = 0.65$), suggesting that any negative effects induced by the research wore off over time, $F(1, 47) = 7.09$, $p = .01$. Most important, awareness of whether they had been deceived had no effect on participants' affective response ($p > .8$, $\eta^2 = .001$), suggesting that participants were more disturbed by negative feedback than by their awareness of deception. None of the interaction effects in this analysis were significant (all $ps > .5$).

Suspicion

One of the suspicion items asked participants about future suspicion regarding any information presented by an experimenter. Lower scores on this item indicate greater future suspicion. Participants who were given a full debriefing reported more future suspicion ($M = 3.63$, $SD = 1.74$) than did participants who were given a partial debriefing ($M = 4.75$, $SD = 1.67$), $F(1, 47) = 7.03$, $p = .01$. This suspicion increased slightly over time across all participants (suspicion at the first session $M = 4.42$, $SD = 1.77$; at the third session $M = 3.96$, $SD = 1.80$), $F(1, 47) = 3.43$, $p = .07$. However, this increase is most likely an artifact of the experimental design, as the interaction effect described later in this article makes clear.

Participants who were given a partial debriefing initially reported little suspicion ($M = 5.27$, $SD = 1.48$). After 3 months (and a full debriefing in the second session), their suspicion increased ($M = 4.21$, $SD = 1.79$) to levels more similar to those reported by participants who were given a full debriefing (M for first session = 3.58, $SD = 1.64$; M for 3 months = 3.72, $SD = 1.79$), $F(1, 47) = 9.63$, $p < .05$. This is not surprising because both groups were

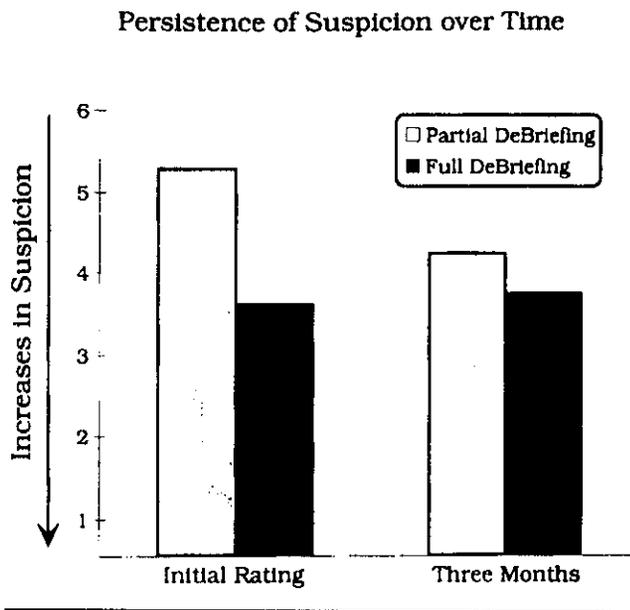


Figure 1 The persistence of suspicion over time, shown as an interaction of time of measurement with debriefing condition.

now aware of the deception, which was not the case after the first session. Note that these latter three cell means fall on or below the scale midpoint. This convergence suggests, as Kelman (1967) had feared, that participants' knowledge of deception in psychology research may in fact create more suspicion of psychological research and that this suspicion is maintained for at least 3 months after the experience.

The second suspicion item asked participants the extent to which they felt psychologists were trustworthy and honest. This was more a measure of suspicions about psychologists than a measure of suspicions about experiments. Participants' responses to this question hovered about 1 point above the scale midpoint ($M = 5.31$, $SD = 1.26$) and were thus mildly positive. There were no experimental effects on this measure (all $ps > .05$).

Educational Benefit

Self-reported benefit. In response to our question about whether they had received significant educational benefit, participants felt that there was moderate benefit ($M = 5.32$, $SD = 1.03$). There were no significant main effects or interactions involving this measure (all $ps > .2$). This contradicts previous research that suggests that participants in deception experiments report more educational benefit than participants in nondeception experiments (Smith & Richardson, 1983).

Free and cued recall. Two other objective measures of educational benefit were participants' free recall about the experiment during the second session and their answers to the 16 multiple-choice questions adminis-

tered at the same time. A 2 (debriefing: partial, full) \times 2 (performance feedback: positive, negative) ANOVA conducted on the total number of items participants recalled yielded no significant main effects or interactions for either of the memory measures. However, these null effects are difficult to interpret given the poor reliability and questionable validity of these measures.

DISCUSSION

In this experiment, participants' awareness of deception was manipulated between conditions and thus provided a relatively clean look at responses to deception, uncomplicated by most confounding factors. Some research (see Christensen, 1988) suggests that deception actually has little negative effect on participants and that the presence of deception may even increase their level of enjoyment in research (Smith & Richardson, 1983). The present research makes clear the complications in this story. Our participants' affective responses did not vary depending on whether they were aware or unaware of the deception. Participants did, however, report negative effects from receiving negative performance feedback, although the effects appeared to be alleviated considerably over time. In addition, certain kinds of suspicion were increased by the awareness of deception. These increases remained for at least the 3-month interval investigated in this experiment. Finally, awareness of deception did not seem to increase claims for educational benefit, and objective measures of memory did not show any effects. The final story, then, is more complex than the two claims often made, namely, that participants enjoy and learn from deception or that participants are clearly harmed by deception.

Deception Versus Other Negative Stimuli

Participants in our experiment were more adversely affected by receiving negative performance feedback than by being deceived. Mean scores on the affective response index were nearly identical across debriefing conditions (5.02 and 4.98), but there was a small overall effect of about 1/2 of a scale point for affective response as a function of receiving negative or positive feedback across all sessions ($\eta^2 = .093$). Thus, to the extent that participation in this study produced negative affect in participants, it was a function of receiving negative performance feedback, not of being deceived.

Psychologists employ a variety of techniques that might create some discomfort or negative affect for participants. Most ink, however, has been spilled in debates over the propriety of deception. Our results suggest that other techniques, such as negative performance feedback and other self-involving manipulations, are likely to have greater effects than deception on participants' affective responses. There seems little warrant for

concern that deception harms participants in any way they notice. It may, in fact, be an issue of participants' rights or experimenters' duties that keeps the issue of deception alive in discussions of experimental ethics. If this is the case, these disagreements are unlikely to be resolved by any empirical work.

It should be noted that affective responses in this experiment became more positive over time, although the absence of a significant interaction suggests that the gap between participants receiving positive versus negative feedback was not significantly narrowed. It seems unlikely that an experimenter could give negative performance feedback to participants without having some effect on their affective reactions. Even if this were possible, the point of these manipulations is usually to produce a motivational state, which would likely be lost if the participants were unfazed. Thus, for some research it is inevitable that participants will experience negative affect or at least decreases in positive affect. However, our results suggest that the return to normal affective states may be more difficult to achieve than previously thought.

Educational Benefit

Other investigators (Smith & Richardson, 1983) have suggested that participants may actually receive more educational benefit from participation in deception research. In this experiment, however, no evidence of increased educational benefit was found in participants' self-reports. Furthermore, no significant differences were discovered in participants' memory for relevant information across debriefing conditions. The self-report and free-recall data seem relatively clear and unambiguous. The low reliability of the multiple-choice items casts some doubt on their worth as a meaningful measure. But, at minimum, a reasonable effort was made to identify educational benefits that arise from an awareness of deception, and such benefits proved elusive.

These findings suggest that Smith and Richardson's (1983) results were either the product of biased self-reports or differential debriefings but not of actual educational benefit resulting from an awareness of deception manipulations (via dissonance reduction or other avenues). Consequently, it seems plausible that different levels of reported educational benefit reflect either the types of debriefings presented in deception versus non-deception experiments (Coulter, 1986) or are related to some other confound, such as more interesting research situations or increased personal attention from experimenters. However, given the weakness of our objective measure of educational benefit, this proposition is only tentative. Further research using more valid measures of educational benefit will be required to make any statements unequivocally. Additional research may be impor-

tant for two reasons. First, a good measure of educational benefit from experimental participation would help resolve the now problematic claim that deception research can be justified because of the additional educational benefit that participants receive. Second, many required research participation pools are founded on the principle that educational benefit is received from research participation, and it seems reasonable to ask that some evidence for this claim be offered.

Suspicion

Finally, Kelman's (1967) fears that deception research will increase levels of suspicion among potential participants were partially confirmed. Differences in suspicion about experiments were found, but these were not accompanied by differences in the belief that psychologists are trustworthy and honest.

Participants who were aware of the deception in our experiment reported relatively high levels of suspicion of experiments after the first session, whereas participants who were not aware of the deception reported little suspicion. In the 3-month follow-up interview, after everyone was informed of the deception, relatively high levels of suspicion of experiments were reported in both groups. Thus, awareness of deception did create suspicion among participants that remained over time. This effect was complicated, however, by the lack of any changes in suspicion about whether psychologists in general are trustworthy and honest. The lack of effects on this more character-based measure may be simple demand (e.g., it is difficult to say "psychologists are liars" to an experimenter over the telephone) or it may suggest that participants were unwilling to generalize to all psychologists on the basis of their limited experience in an experiment in which the psychologists admitted to being constrained by experimental necessity.

There seems little evidence that the suspicions of participants are contagious to so-called naive participants, at least in this campus environment. Our participants reported little suspicion about experiments unless informed of the deception. Thus, they did not enter the lab with heightened suspicion, even though deception experiments are commonly conducted at our institution. This is consistent with other suspicion of research (Sharpe et al., 1992) and suggests that fears about highly wary subject pools may be unwarranted. Undoubtedly, reliance on this effect should be tempered by the realization that it is probably the low profile of these experiments on campus that reduces contagion. If a spectacular instance of deception were to appear in campus newspapers, for example, then contagion might be a serious problem.

In addition, some researchers have recently suggested that deception compromises the reputation of psycho-

logical experimenters among potential participants (Ortmann & Hertwig, 1997). In turn, this may undermine the cooperative principle between participant and experimenter, thereby transforming "every interaction between an experimenter and a participant into a repeated game" (Ortmann & Hertwig, 1997, p. 747). These reputational spillover effects could potentially contribute to high variability and nonreplicability in psychology experiments. Although our data do not investigate this possibility directly, we believe that this effect is considerably more complicated. Given the lack of suspicion among our uninformed participants, spillover from other students or the creation of a general, negative stereotype seems unlikely. Furthermore, our participants were not disturbed by the deception, and it seems unlikely that any negative impressions or later discussions would occur due to a manipulation that was perceived uneventfully. However, the stability of suspicion following our experiment makes it unclear whether our participants would behave differently if they participated in subsequent psychological experiments. Thus, the negative effects of suspicion seem to occur within subjects, not between. Research using behavioral measures of suspicion levels over time among naive participants would be necessary to determine the effects of any increased suspicion on actual behavior in psychological experiments.

CONCLUSION

Although there is no evidence here to support the claim that participants enjoy deception experiments more than nondeception experiments, our participants did not perceive deception in psychological research as a source of negative feelings and emotions. When our participants discovered that they had been deceived, they perceived no additional direct costs in terms of their affective reactions. Those who were responsible for cost-benefit analyses of psychological experiments may now have sufficient evidence to lower the weighting of mere deception as a cost to participants.

Participants' reactions of suspicion and their affective reactions to negative, false feedback do not lead to such a sanguine view. There is sufficient evidence of participants' suspicion about experiments to cause concern among researchers. Suspicion of experiments increased after learning about deception and remained high after an interval of several months. Even if this suspicion does not generalize to all psychologists, further investigation of the effects of this suspicion on later participation in research seems warranted. Finally, the potentially nonintuitive nature of these results suggests that we should turn our ethical attention to the that costs our *participants* perceive rather than to those we impute to them.

NOTES

1. In the literature, deception is normally defined as any purposeful attempt to "actively mislead participants regarding some aspect of the study" (Adair, Dushenko, & Lindsay, 1985, pp. 62-63). Withholding information or failing to disclose the true purpose of a study is not commonly counted as deception. We have chosen to use this definition throughout our article as well.

2. The research participation pool at St. Olaf College is designed to allow students who object to participating in deceptive research to avoid it. Each term, only a small handful of students (less than 5 from the pool of 200 to 250 students) select this option. These individuals are personally notified of a subset of nondeceptive studies for which they may sign up. The remaining participants thus agree, at least implicitly, to potentially experience deception. In the term in which this experiment was conducted, no students opted out of deception experiments, and so the representativeness of our sample was not compromised.

3. Taylor and Shepperd (1996) have argued that probes for suspicion may be ineffective because subjects may be unwilling to report suspicion. They report an episode in which participants uncovered deceptive procedures after discussing the experiment with one another, despite specific instructions to refrain from discussion. Yet, when later asked about suspicion, none of these participants reported any, even after careful questioning. We suspect Taylor and Shepperd's participants were unwilling to voice their suspicions because they had seriously violated the experimenter's instructions. During pilot testing for the current experiment, numerous participants reported being suspicious of a more severe feedback manipulation (i.e., bottom 20%). Given this willingness to report suspicion during the pilot testing, we believe that the probes were valid and that no subjects were actually suspicious of our methods.

REFERENCES

- Adair, J. G., Dushenko, T. W., & Lindsay, R.C.L. (1985). Ethical regulations and their impact on research practice. *American Psychologist, 40*, 59-72.
- Aitkenhead, M., & Dordoy, J. (1985). What the subjects have to say. *British Journal of Social Psychology, 24*, 293-305.
- Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and non-depressed students: Sadder but wiser? *Journal of Experimental Psychology: General, 108*, 441-485.
- Alloy, L. B., & Abramson, L. Y. (1982). Learned helplessness, depression, and the illusion of control. *Journal of Personality and Social Psychology, 42*, 1114-1126.
- American Psychological Association (APA), Committee for the Protection of Human Participants in Research (1982). *Ethical principles in the conduct of research with human participants*. Washington, DC: Author.
- Arkin, R., & Maruyama, G. (1979). Attribution, affect, and college exam performance. *Journal of Educational Psychology, 71*, 85-93.
- Aronson, E., & Carlsmith, J. M. (1963). Effect of severity of threat on the valuation of forbidden behavior. *Journal of Abnormal and Social Psychology, 66*, 584-588.
- Aronson, E., Ellsworth, P., Carlsmith, J. M., & Gonzales, M. H. (1990). *Methods of research in social psychology* (2nd ed.). New York: McGraw-Hill.
- Baumrind, D. (1985). Research using intentional deception: Ethical issues revisited. *American Psychologist, 40*, 165-174.
- Berscheid, E., Baron, R. S., Dermer, M., & Libman, M. (1973). Anticipating informed consent: An empirical approach. *American Psychologist, 28*, 913-925.
- Christensen, L. (1988). Deception in psychological research: When is its use justified? *Personality and Social Psychology Bulletin, 14*, 664-675.
- Collins, F. L., Jr., Kuhn, I. F., Jr., & King, G. D. (1979). Variables affecting subjects' ethical ratings of proposed experiments. *Psychological Reports, 44*, 155-164.
- Coulter, X. (1986). Academic value of research participation by undergraduates. *American Psychologist, 41*, 317.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

- Fisher, C., & Fyrberg, D. (1994). College students weigh the costs and benefits of deceptive research. *American Psychologist*, *49*, 417-427.
- Fleming, M., Bruno, M., Barry, K., & Fost, N. (1989). Informed consent, deception, and the use of disguised alcohol questionnaires. *American Journal of Drug and Alcohol Abuse*, *15*, 309-319.
- Gerdes, E. P. (1979). College students reactions to social psychological experiments involving deception. *Journal of Social Psychology*, *107*, 99-110.
- Kelman, H. C. (1967). Human use of human subjects: The problem of deception in social psychological experiments. *Psychological Bulletin*, *27*, 1-11.
- Korn, J. (1987). Judgments of acceptability of deception in psychological research. *Journal of General Psychology*, *114*, 205-216.
- Krupat, E., & Garonzik, R. (1994). Subjects expectations and the search for alternatives to deception in social psychology. *British Journal of Social Psychology*, *33*, 211-222.
- MacCoun, R. J., & Kerr, N. L. (1987). Suspicion in the psychology laboratory: Kelman's prophecy revisited. *American Psychologist*, *42*, 199.
- McNamara, J. R., & Woods, K. M. (1977). Ethical considerations in psychological research: A comparative review. *Behavior Therapy*, *8*, 703-708.
- Milgram, S. (1964). Issues in the study of obedience: A reply to Baumrind. *American Psychologist*, *19*, 848-852.
- Oliansky, A. (1991). A confederate's perspective on deception. *Ethics and Behavior*, *1*, 253-258.
- Ortmann, A., & Hertwig, R. (1997). Is deception acceptable? *American Psychologist*, *52*, 746-747.
- Pettit, P. (1991). Consequentialism. In P. Singer (Ed.), *A companion to ethics* (pp. 230-240). Cambridge, MA: Blackwell.
- Pihl, R., Zaccchia, C., & Zeichner, A. (1981). Follow-up analysis of the use of deception and aversive contingencies in psychological experiments. *Psychological Reports*, *48*, 927-930.
- Rubin, Z. (1985). Deceiving ourselves about deception: Comment on Smith and Richardson's "Amelioration of deception and harm in psychological research." *Journal of Personality and Social Psychology*, *48*, 252-253.
- Schwartz, S., & Gotlieb, A. (1981). Participants' post-experimental reactions and the ethics of bystander research. *Journal of Experimental Psychology*, *17*, 396-407.
- Seemans, J. (1969). Deception in psychological research. *American Psychologist*, *24*, 1025-1028.
- Sharpe, D., Adair, J., & Roese, N. (1992). 20 years of deception research: A decline in subjects' trust? *Personality and Social Psychology Bulletin*, *18*(5), 585-590.
- Sherer, M., Maddox, J. E., Mercandante, B., Prenice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The Self-Efficacy scale: Construction and validation. *Psychological Reports*, *51*, 663-671.
- Smith, C. P. (1981). How (un)acceptable is research involving deception? *IRB: A Review of Human Subjects Research*, *3*(8), 1-4.
- Smith, C. P., & Richardson, D. (1983). Amelioration of deception and harm in psychological research: The important role of debriefing. *Journal of Personality and Social Psychology*, *44*, 1075-1082.
- Taylor, K. M., & Shepperd, J. A. (1996). Probing suspicion among participants in deception research. *American Psychologist*, *51*, 886-887.

Received May 28, 1996

Revision accepted August 4, 1997