Failure to Recognize the Effect of Implicit Social Influence on the Presentation of Self

Jacquie D. Vorauer
Department of Psychology University of Manitoba

Dale T. Miller
Department of Psychology Princeton University

ABSTRACT

Two studies demonstrated that individuals can fail to detect changes in their actions that are induced by implicit social influence. In both studies, observers' impressions indicated that actors matched the positivity of their remarks about themselves to the positivity of another person's self-description. However, actors' own judgments of the types of impressions they conveyed revealed that they did not perceive the effect of the other's self-description on their self-presentation. Study 1 suggested that actors' relatively poor access to their own nonverbal behavior could not fully account for their failure to perceive how they were influenced. Study 2 indicated that actors' metaperceptions were connected to actors' general beliefs about themselves, whereas observers' impressions were not. The "blindness" effect was driven primarily by actors low in self-esteem. Implications for self-presentation and other social phenomena are discussed.
concerns with justice or equity (Jackson & Harkins, 1985), the potential informational value of other people's actions (Rohrer, Baron, Hoffman, & Swander, 1954), and schema activation (Carver, Ganellen, Froming, & Chambers, 1983) might contribute to individuals' susceptibility to other people's actions. However, despite this intensive examination of social influence processes and a general appreciation of their potential consequences for decision making and interpersonal relations, the research on this topic has been unbalanced. The important issue of how the targets of social influence perceive their own responses to it has gone largely unaddressed. How sensitive are people to the fact that their actions shift in accordance with the actions of others? Are individuals apt to recognize when another person's behavior prompts them to behave more competitively than they originally intended, to sound more confident than they really feel, or to express more liberal attitudes than they truly possess?

### Implicit Assumption of Awareness

To date, when self-inferences have been considered in the context of social influence, the focus has been on individuals' attributions for behavior induced by an interaction partner rather than on whether they even recognize that the behavior has occurred (e.g., Darley & Fazio, 1980; Kelley & Stahelski, 1970; Snyder & Swann, 1978). Investigators have not directly examined people's perceptions of their current actions or the possibility that people might fail to discern changes that are apparent to an outside observer. The neglect of this issue in the literature to some extent seems to reflect an implicit assumption that people are aware of when their current actions deviate from what they had intended or are not representative of their usual behavior. An assumption of awareness would seem reasonable in experimental contexts in which participants' behavior consists of naming the longest line in a set (Asch, 1951), choosing to either cooperate or defect (Kelley & Stahelski, 1970), or deciding on the level of a noise weapon to administer to someone else (Snyder & Swann, 1978). Under such circumstances, there is little subjective judgment involved in characterizing the person's behavior or the extent to which it is consistent with the behavior of his or her interaction partners. Consider, however, the much more flexible behavioral repertoire that is typically available to individuals in their everyday social exchanges. For example, people's accommodations to another person might be manifest in their expressing a select subset of their personal beliefs about an issue, emphasizing their feelings about themselves in one domain rather than other domains, or altering nuances of their actions and nonverbal behaviors (e.g., more qualified and less enthusiastic statements of their opinions). In cases such as this, in which judgments about the nature of behavior are more subjective, it seems questionable whether individuals will recognize how they have been affected by social forces, particularly if those forces are not of an explicit nature.

In the present research, we sought to illustrate that people can be insensitive to the extent to which they engage in behavior matching. Our primary goal was to demonstrate that individuals may fail to detect changes in their actions that are induced by implicit social influence, changes that are reflected in observers' impressions. Note, then, that we were not concerned with examining individuals' introspective access into the factors causing them to behave in a particular way (Nisbett & Wilson, 1977). Our concern was more basic, focusing simply on people's ability to recognize whether they were behaving one way as opposed to another. However, our hypothesis does represent an extension of research indicating that individuals' causal accounts for their behavior may neglect important social influences (LaFtane & Darley, 1968). Our prediction suggests that such oversights can be evident in perceptions of behavior as well as in explanations for it.

In our investigations, the impressions formed by observers provided the necessary standard of comparison for actors' perceptions. Actors were asked to estimate how an observer who had access only to their current behavior would perceive them, and these metaperceptions (Kenny & DePaulo, 1993) were compared to the observer's actual impressions. Observers' judgments were taken as the criterion not because observers' inferences are necessarily more correct or valid than actors' inferences, but because we were particularly interested in individuals' ability to assess their behavior as it appears to others. The extensive literature on self-presentational phenomena (Baumeister, 1982; Goffman, 1959; Jones, 1964; Schlenker, 1980) has highlighted the importance people accord to others' impressions of them and has suggested that individuals are often concerned about how they appear to others during their everyday social interactions (see also Sheldon & Johnson, 1993).

### Potential Accounts for Actors' Insensitivity

Why might actors fail to perceive the effects of current circumstances on their actions? We identified two key possible contributing factors, either or both of which might account for actors' insensitivity to how they engage in behavior matching. First, actors' inaccurate metaperceptions may reflect their limited access to their own nonverbal behaviors. Many of individuals' responses to their social context take place in their body movements and facial expressions (e.g.,
Swann, Stein-Seroussi, & McNulty, 1992; Word et al., 1974). Research has indicated that even "small slices" of such nonverbal activity can contain significant information (Ambady & Rosenthal, 1992, 1993). Actors, however, cannot see how these behaviors appear (physically) to observers (Barr & Kleck, 1995; DePaulo, 1992; Jones & Nisbett, 1972). Discrepancies between individuals' metaperceptions and others' actual impressions may thus stem from critical information being unavailable to actors, or at least less available to them than it is to observers. For example, individuals may not realize how their facial expressions shift to convey more or less warmth, self-confidence, or agreement with their interaction partners.

Second, actors might be essentially "blinded" by their wealth of self-knowledge. The extensive information that individuals possess about their general personal qualities as well as about their feelings in the moment may make it difficult for them to perceive any outward deviations from their "true" self. There are a number of processes through which such blinding could occur. One possibility is that the availability of general self-knowledge leads individuals to ignore their current behavior altogether. Consider, for example, the case in which individuals are induced to communicate a biased subset of their experiences or feelings. Depending on the remarks of their interaction partner, people might emphasize the negative aspects of their relationship with their child or the positive qualities of their romantic partner. If individuals assess the impression they conveyed by referring only to their personal knowledge, they may fail to recognize that they have communicated a select sample of their experiences that is not actually representative of their overall feelings about the person or relationship.

Another possibility is that people consider their current behavior but interpret it in a biased manner. They may focus selectively on actions and remarks that are consistent with their private self-knowledge and feelings, attach (self-congruent) meaning to ambiguous behaviors, or both. Numerous investigations and theoretical analyses suggest that individuals' self-knowledge colors how they perceive their own actions (Kenny & DePaulo, 1993; Markus & Sentis, 1982). Consider, for example, a group discussion in which individuals are led to acknowledge valid points associated with a particular dominant position. People might perceive that their inner reservations were readily apparent in questions they asked, the points at which they were silent or hesitant, and the fact that they refrained from explicitly endorsing the position. An observer, however, may interpret these subtleties differently and view individuals' acknowledgment of valid points as indicative of whole-hearted agreement with the position. Thus, even though a person's self-concept might generally provide a good estimate of the impressions he or she conveys, in the presence of an impactful external influence on behavior reliance on such self-knowledge may produce inaccurate inferences.

Note that for possession of self-relevant information to account for people's insensitivity to their own behavior matching, self-knowledge must do more than guide their perceptions. It must also reduce individuals' propensity to consider other relevant sources of information or perspectives on their behavior.

Overview

In both of the present studies, we sought to demonstrate that actors can fail to fully appreciate how their behavior is affected by implicit social influence. We also began to explore the role played by each of the two potential contributing factors that we have identified. In Study 1, we examined whether actors' relatively poor access to their own nonverbal behavior might account for their insensitivity to how they engage in behavior matching. Next, in Study 2, we examined the possible contribution of actors' private self-knowledge to their blindness relative to observers. We did not necessarily expect that actors would fail to detect any effect of implicit social influence on their actions. However, we did anticipate that the effects perceived by actors would be significantly smaller than those indicated by observers' impressions. Thus, by the terms insensitivity or blindness we mean to capture this relative actor—observer difference rather than to imply a complete failure by actors to recognize the impact of implicit social influence on their behavior.

It is interesting to juxtapose our proposal regarding the blinding effect of self-knowledge with Bem's (1972) analysis of self-perception processes. Bem maintained that when internal cues are weak or ambiguous, individuals are functionally in the same position as an outside observer: They come to know their own attitudes and emotions by drawing inferences from their behavior. Our main thesis is parallel but opposite. We suggest that when internal cues are strong and unambiguous, individuals do not see their behavior as it appears to an outside observer because they come to know their own behavior in part by inferring it from internal self-knowledge.

Study 1

Study 1 was designed to provide a test of our basic hypothesis regarding actors' and observers' differential sensitivity to
behavior-matching processes. We developed a paradigm that would allow us to expose actors to social influence and examine its effects on their behavior versus their metaperceptions. It was important that the social influence be relatively implicit in nature and that actors' response options be varied and not easily quantifiable. These conditions would capture important elements of everyday social interactions.

The investigation focused on behavior matching with respect to the positivity of self-presentations. Pairs of same-sex unacquainted students were recruited for a study of "undergraduates' academic and social experiences." Pair members were randomly assigned to either the actor or observer condition. Actors were interviewed about their experiences by their observer partner, who asked a series of predetermined questions. These interviews were videotaped. Immediately prior to being interviewed by their observer partner, actors read a questionnaire that he or she had ostensibly completed. The questionnaire described either positive experiences and high self-esteem or negative experiences and low self-esteem. The interview questions asked actors to describe positive and negative aspects of their academic and social experiences in an open-ended manner. After the interview, actors estimated the impressions they had conveyed along a number of dimensions (e.g., self-confidence, satisfaction with academic accomplishments and social life); observers indicated the impressions they had actually formed. Actors and observers also estimated the percentage of time that actors spent discussing positive versus negative experiences.

We anticipated that actors' self-presentations would vary according to whether they read a positive or negative description prior to their own interview: Observers' judgments should reveal more positive impressions of actors in the positive stimulus condition than of actors in the negative stimulus condition. Our main hypothesis was that actors would perceive a smaller shift in their actions and remarks than would be indicated by observers' judgments: Actors' metaperceptions should be relatively similar across the positive and negative stimulus conditions. Such divergent effects for actors versus observers (i.e., metaperceptions vs. behavior) would be reflected by a significant Actor—Observer x Positive—Negative Stimulus interaction effect. We expected parallel results on participants' estimates of the percentage of time that actors spent discussing negative versus positive experiences.

There were two additional sources of data. We asked actors to indicate their current beliefs about their true self, as an indirect means of assessing the extent to which they perceived any discrepancy between their self-presentations and their actual inner feelings. We did not expect that actors would sense any systematic discrepancy; if they did, we anticipated that it would be constant across the positive and negative stimulus conditions. Finally, we videotaped and transcribed the interviews so that they could be coded by judges. If the expected pattern of divergent effects on actors' metaperceptions and observers' impressions was obtained, coding the transcripts would allow us to explore whether there was sufficient information in the content of actors' speech for the impact of the stimulus manipulation to be detected. An effect for the stimulus manipulation on the content of actors' speech would detract from an explanation of actor—observer discrepancies solely in terms of these individuals' differential access to important information (actors know what they said, if not how they looked) and would suggest a significant role for other factors such as actors' overreliance on self-knowledge.

Method Participants

Seventy Princeton University undergraduates (38 women and 32 men) were paid $5 for their participation in the 1-hr. experiment. Approximately half of the participants were recruited from sign-up sheets posted in the psychology building; the remainder had indicated at registration an interest in participating in psychology studies for payment, and we contacted these participants by telephone. Participants were scheduled in same-sex pairs. Prior to each session it was ascertained that pair members were unacquainted with one another. There were 17 (8 male and 9 female) pairs in the negative stimulus condition, and 18 (8 male and 10 female) pairs in the positive stimulus condition.

Procedure

On their arrival at the laboratory, participants were instructed that the study was an investigation of undergraduates' experiences at Princeton. Ostensibly, the researchers were interested in students' feelings about different aspects of life on campus. The experimenter explained that different methods of asking students questions about their experiences were being used, so that the response formats that students were most comfortable with and the methods that yielded the best information could be identified.

Assignment to actor and observer roles.
The experimenter told participants that the two response options in the present study involved either filling out a questionnaire or answering the same questions out loud in an interview format. She flipped a coin in front of participants to randomly assign one of them to each method. Unbeknownst to participants, the coin flip simultaneously served to determine their assignment to actor and observer roles in the study. The interview participant would be the actor, and the questionnaire participant would be the observer. Note that although we use the terms actor and observer in our description of the procedure, the experimenter did not use these labels in her communications with participants.

**Experimental manipulation.**

The experimenter explained that the next step in the study was for the observer to complete the questionnaire (this procedure was included only to enhance the plausibility of the cover story). She advised participants that while the observer filled out the questionnaire, the actor could look over the questions and organize his or her thoughts for the upcoming interview. At this point, the observer was escorted to a separate room. After 15 min, the experimenter brought the actor a questionnaire that had ostensibly been filled out by the observer. In fact, the actor was given a preconstructed questionnaire that varied according to his or her experimental condition. Actors in the positive stimulus condition received a questionnaire that described positive academic and social experiences that were accompanied by feelings of high self-esteem and mastery. Actors in the negative stimulus condition received a questionnaire that described negative academic and social experiences that were accompanied by feelings of low self-esteem and loneliness. The contents of the female versions of each of these questionnaires are presented in the Appendix (men received parallel versions). At this time, actors were also given a two-item survey to complete. They were instructed to read over the observer's questionnaire and to rate the positive and the negative experiences that he or she described in terms of how typical versus unusual they were for a Princeton student. This task was designed to provide actors with a rationale for reading the observer's questionnaire. When actors were given the questionnaire to read, they were told that the observer had consented to this procedure. They were instructed, however, to avoid referring to the content of the observer's description in their own interview, which was to focus solely on their own personal experiences. After actors had read the questionnaire and made their ratings on the survey, the experimenter instructed the observer that the time limit for completing the questionnaire had been reached. He or she was then brought back into the room with the actor for the interview.

**The interview.**

To control the potential influence of social feedback on actors' metaperceptions, the observer's role in the interview was highly structured and constrained. However, so that the observer's task would appear more sensible, participants were led to believe that there would be a subsequent component of the interview in which the observer's involvement would be more significant (this second part of the interview never took place). The observer followed a script that prompted the actor to describe two general categories of experiences. These two general questions were identical to those that appeared on the questionnaire completed earlier by the observer and given to actors for their perusal prior to the interview. The observer first provided the actor with the following instructions:

Please take a moment to think about your academic accomplishments at Princeton. First describe some positive things that come to mind in this regard, then discuss any negative experiences you have had. Please also describe the strength of the impact each experience had on you (e.g., your feelings about yourself).

After the actor completed describing these experiences, the observer instructed him or her to "now please think about the people you have met at Princeton and the social activities in which you have participated." The remaining instructions for the social experiences component of the interview were the same as those described above for the academic experiences component.

Actors decided for themselves when to move from discussing positive experiences to discussing negative experiences. The experimenter informed actors at the outset that if they spent longer than 5 min on one of the two main sections of the interview, she would interrupt and ask them to finish. This happened on only one occasion. The interview was videotaped; the camera was placed to the side but focused solely on the actor. Actors were instructed that the interview was to be a "videotaped version of a questionnaire."

**Dependent Measures**

Participants were first asked a filler question about their preference for different response formats. Actors then indicated the impressions they thought that their answers to the interview questions conveyed about themselves (i.e., their
These data were collected by asking a question that focused directly on their own perception of their behavior: "Approximately what percentage of the time you spoke about academic experiences did you speak about positive versus negative experiences?" They divided 100% between positive and negative experiences. They then responded to a similar question about their social experiences. Observers answered parallel questions, estimating the percentage of time the actor spent discussing positive versus negative academic and social experiences.

Results

Preliminary analyses revealed no significant effects for gender when it was included as a factor. This variable is therefore excluded from the analyses reported below.

Manipulation Check

The impressions that actors in the positive and negative stimulus conditions formed of the observer on the basis of the preconstructed questionnaires were compared. A t test revealed that the manipulation was effective. Actors in the positive condition perceived the observer as more satisfied with his or her self and university experiences ($M = 4.32$) than did actors in the negative condition ($M = 4.29$), $t(33) = 10.11, p < .001$.

Satisfaction: Metaperceptions Versus Impressions

To assess whether actors' metaperceptions and observers' impressions were differentially affected by the stimulus manipulation, participants' responses along the five satisfaction dimensions were combined to form indexes. These data were analyzed in a 2 (actor vs. observer) × 2 (positive vs. negative stimulus) repeated measures analysis of variance (ANOVA), with pairs as the unit of analysis. The first factor was a within-pairs variable, and the second factor was a between-pairs variable. This analysis yielded an Actor-Observer × Stimulus interaction, $F(1, 33) = 5.43, p = .025$. The first question that we examined in our simple effects analyses centered on whether actors did in fact engage in behavior matching: Did they alter their own self-presentations in light of the experiences (ostensibly) shared by the other person? The results for observers' judgments confirmed that actors conveyed significantly more positive impressions in the positive stimulus condition ($M = 4.63$) than in the negative stimulus condition ($M = 4.12$), $F(1, 57) = 5.43, p = .025$. Next, we examined the central question of whether actors were unaware of the extent to which they had engaged in behavior matching: Did actors perceive a smaller behavioral effect than was indicated by observers' judgments? The analysis revealed that actors perceived no effect whatsoever of the stimulus manipulation on their own self-presentations. Actors' metaperceptions were virtually identical across the positive ($M = 4.29$) and negative ($M = 4.32$) stimulus conditions, $F(1, 57) < 1$. The results thus confirmed our prediction that actors would fail to appreciate the extent to which their self-presentations were affected by their current social context. An alternative way of examining the interaction is to compare actors' metaperceptions to observers' impressions in each of the two stimulus conditions. These analyses indicated that in the negative stimulus condition, actors' metaperceptions were not different from observers' impressions, ($M = 4.32$ and 4.12, respectively), $F(1, 33) = 1.26, ns$. In the positive stimulus condition, actors' metaperceptions were nonsignificantly less positive ($M = 4.29$) than observers' impressions ($M = 4.63$), $F(1, 33) = 3.96, p = .056$. There were no other effects. To examine the impact of the experimental manipulation more precisely, we computed the effect size for the influence of the stimulus factor on observers' impressions; in doing so, we implemented a correction recommended by Hedges and Olkin (1985) to take into account overestimation of population effect sizes from small samples. The resulting value was .76. In light of Cohen's (1977) advice that effect sizes of .20 be regarded as small, .50 as medium, and .80 as large, the effect size obtained in the present study was quite impressive. Finally, although actors and observers were differentially affected by the stimulus manipulation, the average within-cell correlation between actors' metaperceptions...
and observers' impressions was significant, \( r (29) = .38, p < .05 \).

**Perceptions of Behavior**

We expected that participants' perceptions of the percentage of time the actor spent discussing positive versus negative experiences would follow the same pattern as their judgments on the satisfaction measure. That is, we hypothesized that observers would perceive a greater emphasis on positive experiences in the positive as compared with the negative stimulus condition, and that actors' own judgments would be insensitive to the magnitude—or perhaps the existence—of this effect. To test this prediction, participants' estimates for academic and social experiences were combined and analyzed in a 2 (actor vs. observer) \( \times \) 2 (positive vs. negative stimulus) repeated measures ANOVA, with pairs as the unit of analysis.

The first factor was a within-pairs variable, and the second was a between-pairs variable. This analysis yielded anActor-Observer \( \times \) Stimulus interaction, \( F (1, 33) = 5.59, p < .025 \). The pattern of results confirmed our hypothesis. As expected, observers' judgments indicated behavior matching by actors: Observers perceived that actors in the positive condition focused on positive experiences to a greater extent (\( M = 58.06 \)) than did actors in the negative condition (\( M = 49.85 \)), \( F (1, 47) = 3.91, p = .05 \). Once again, however, actors' own perceptions of their behavior were remarkably similar across the positive and negative stimulus conditions (\( M = 55.56 \) and 55.15, respectively), \( F (1, 47) < 1 \). An alternative way of examining this interaction is to compare actors' perceptions to observers' perceptions in each of the stimulus conditions. These analyses revealed that in the negative condition, observers perceived that actors placed less emphasis on positive experiences than the actors themselves believed, \( F (1, 33) = 5.01, p < .05 \). The perceptions of actors and observers in the positive condition were not significantly different, \( F (1, 33) = 1.18, ns \). Thus the overall pattern obtained for perceptions of behavior was very similar to that obtained on the satisfaction measure: Observers' perceptions were affected by the stimulus manipulation, whereas actors' perceptions were not. On this measure, however, the actor—observer discrepancy occurred in the negative stimulus condition rather than in the positive condition. There were no other effects. The average within-cell correlation between actors' and observers' perceptions of behavior was high, \( r (29) = .68, p < .001 \).

**Actors' Self-Ratings**

To examine whether actors felt that there was a discrepancy between their true self and the self that they presented, and whether their sense of discrepancy was affected by the stimulus manipulation, we combined actors' self-ratings along the five satisfaction dimensions and entered these data into a 2 (true self vs. metaperception) \( \times \) 2 (positive vs. negative stimulus) repeated measures ANOVA. The first factor was a within-subjects variable, and the second factor was a between-subjects variable. The analysis revealed that although actors generally felt that their true self was more positive (\( M = 4.64 \)) than the impression they had conveyed (\( M = 4.31 \)), \( F (1, 33) = 7.78, p < .01 \), this effect was not qualified by the stimulus manipulation (interaction \( F < 1 \)). A direct examination of actors' ratings of their true self revealed that actors in the negative and positive stimulus conditions rated themselves almost identically (\( M = 4.60 \) and 4.67, respectively). In general, then, it seems that actors either felt that they had been modest in their self-presentations or that they had been ineffective in communicating the extent to which they were satisfied with their personal accomplishments and university experiences.

**Content of Speech**

Finally, we explored the extent to which the changes in actors' self-presentations were apparent in the content of their speech. Five judges coded transcripts of the interviews, making the same ratings as those made by observers in the original study. Reliability was adequate (\( \alpha = .70 \) and .77 for impressions of satisfaction and percentage ratings, respectively). (After rating the transcripts, judges also coded the actual videotapes; here the results for their impressions of satisfaction and percentage ratings replicated the significant effects obtained for observers in the original study.) Results from a one-way ANOVA conducted on the percentage ratings made from the transcripts indicated that although judges tended to perceive that actors in the positive condition spent a higher percentage of time discussing positive experiences (\( M = 53.53 \)) than did actors in the negative condition (\( M = 47.02 \)), the effect did not reach conventional levels of statistical significance, \( F (1, 33) = 3.03, p = .09 \). However, results from a one-way ANOVA conducted on impressions of satisfaction indicated that solely on the basis of content of speech, judges formed more positive impressions of actors in the positive condition than of actors in the negative condition (\( M = 4.54 \) and 3.99, respectively), \( F (1, 33) = 9.79, p < .005 \). The results were thus inconsistent across the two measures. The findings suggest that nonverbal information may have been important to the ability to detect an effect of the stimulus manipulation on the percentage of time actors spent discussing positive versus negative experiences. In contrast, there was sufficient information in the content of speech alone for judges to discern an effect of the stimulus manipulation on impressions of satisfaction conveyed by actors. This
latter result by no means suggests that actors did not exhibit nonverbal responses relevant to these judgments; it only highlights that there was sufficient information in the channel accessible to actors for the influence of the stimulus on impressions of satisfaction to be identified. Overall, then, the findings suggest that although differential access to nonverbal information may have been important to the discrepancy between actors' and observers' judgments of the types of experiences that were discussed, other factors probably contributed to actor—observer discrepancies in perceptions regarding the impressions of satisfaction that were conveyed.

Discussion

The results of Study 1 provide clear support for our hypothesis that people can fail to detect the extent to which they have been induced to present a biased impression of their feelings and experiences to an interaction partner. Actors who read another person's positive self-description before describing their own experiences conveyed more positive impressions of their satisfaction with their academic accomplishments and university experiences to observers than did actors who first read another person's negative self-description. Actors' own sense of the impressions they had conveyed about their satisfaction did not reflect this effect, however: Actors' metaperceptions did not vary according to whether they had been exposed to a positive or negative self-description prior to their interview. Actors' and observers' judgments about the extent to which the actor had focused on positive as opposed to negative experiences followed a parallel pattern. Observers perceived that actors in the positive stimulus condition spent more time discussing positive experiences than did actors in the negative stimulus condition. The actors themselves, however, did not perceive a difference in the types of experiences discussed in one condition as opposed to the other. Moreover, actors' sense of whether there was a discrepancy between their true self and how they appeared did not vary across the positive and negative stimulus conditions. In sum, although there was clear evidence that actors' behavior was affected by the information to which they were exposed, actors themselves were not sensitive to how the information influenced the impressions that they conveyed.

The data obtained from the coding of the transcripts allowed us to begin to evaluate different potential accounts for the discrepancies between actors' and observers' judgments. Although the results on the percentage ratings suggested that actors' perceptions may have diverged from those of observers in part because of actors' indirect access to the information contained in their own nonverbal behavior, the results for impressions of satisfaction could not be explained in terms of actors' and observers' differential access to nonverbal cues. On the basis of the content of speech alone, observers formed different impressions of actors in the positive as opposed to the negative stimulus conditions. These findings suggest that other factors probably contributed to actors' insensitivity to the impressions conveyed by their remarks during the interview. In Study 2, we examined whether it was possible to account for actor—observer discrepancies in terms of actors' overreliance on their general self-knowledge.

One might argue that the procedure of Study 1 somehow tricked actors into making erroneous judgments by leading them to misperceive the self-confidence and satisfaction of the observer. Perhaps actors made some correction for the observer's positive or negative personal frame of reference for making judgments, or for the observer's understanding of his or her own potential influence on their self-descriptions. We consider this implausible for a number of reasons. First, we did not ask actors to indicate this particular observer's impression of them. Rather, the cover story for the study and the filler questions in the questionnaire were designed to orient actors to the videotaped interview they had provided about their university experiences and thus to a more general audience for their reactions. Moreover, the expected pattern of actor—observer divergence was obtained for the percentage estimates, which did not ask actors to take any outside perspective on their remarks. It is unclear how a correction process could account for the effects on this measure. Although we believe that it is unlikely that correction processes contributed to the results of Study 1, we made a particularly concerted effort in Study 2 to direct actors to the perspective of an outside observer of their videotaped interview.

Study 2

Study 2 had four objectives. First, we sought to replicate the actor—observer discrepancy effect obtained in Study 1, with the instructions to actors revised to direct them more explicitly to the perspective of an outside observer. Second, this study was designed to shed some light on the information that actors did use in forming their metaperceptions. We anticipated that actors' beliefs about the impressions that they had conveyed would be closely associated with their general feelings about themselves and their university experiences, more closely associated than would observers' actual impressions. This pattern is necessary but not sufficient for demonstrating a blinding effect of self-knowledge. We incorporated a pretest into the design so that we could obtain such indirect support for the role of self-knowledge. Participants' general feelings about themselves and their experiences were assessed several months prior to the experimental session so that their relation to actors' metaperceptions and observers' impressions could be examined.
A third goal was to clarify the nature of actors' behavioral response to the stimulus manipulation. Although we have broadly characterized actors' reactions as behavior matching, it is conceivable that their behavior was driven by a desire to protect the self-esteem of the dissatisfied observer by providing similarly negative experiences or by disclosing information of similar intimacy. If actors' reactions do not reflect motivations centering on the other person, but rather how the different versions of the questionnaire primed distinct aspects of their own experiences or set different norms for self-presentation in the experiment, they should be affected by the stimulus manipulation regardless of whether the questionnaire is presented as belonging to their observer partner. In Study 2, we included a condition in which the stimulus questionnaire was presented as belonging to a third party who was not present during the session and who would not have access to actors' videotapes. We expected that actors' responses to the manipulation would not be contingent on the questionnaire belonging to their observer partner. Such a result would confirm that the actor—observer discrepancy pattern explored in the present research is likely to occur across a broad range of social circumstances.

We explored one final issue in Study 2 related to our thesis that self-knowledge can blind individuals by hindering their detection of how they are affected by implicit social influence. It would seem that some types of self-beliefs might be a source of greater interference in this regard than others. In particular, dissatisfaction and low self-esteem might be more powerful than positive beliefs about the self, in the sense that negative beliefs may preclude the recognition of belief-inconsistent behavior to a greater extent than positive beliefs. Individuals with high self-esteem may be open to the possibility that they appear to be disgruntled or disappointed with themselves, perhaps particularly when (as in the present paradigm) they perceive that they are only part-way through an interaction or self-presentation. In contrast, those with low self-esteem may be disinclined to perceive that they appear content or self-assured. Why might this be the case? Recent research and theoretical analyses suggest that the self-presentation motivations of individuals with low self-esteem are apt to focus on self-protection and the avoidance of risk (Baumeister, Tice, & Hutton, 1989; Tice, 1991). For example, because individuals with low self-esteem wish to minimize their chances of being embarrassed, they avoid predicting success for themselves or expressing positive self-evaluations. This allows them to escape the feeling that unrealistic expectations have been created. Whereas the knowledge that people with high self-esteem possess about themselves might allow for a range of different possible self-presentations depending on their current needs and situation, the self-protective concerns of low self-esteem individuals may prompt them to perceive only a narrow range of impressions that they might convey to someone else.

We should note that this final hypothesis was quite tentative. Existing research does not provide clear grounds for expecting low self-esteem individuals to be less accurate in their metaperceptions than high self-esteem individuals. Indeed, some research has been taken to indicate that depressed individuals' self-perceptions are more accurate than those of nondepressed individuals, who are biased in a self-enhancing direction (Lewinsohn, Mischel, Chaplin, & Barton, 1980). Research by Campbell and Fehr (1990) demonstrated, however, that accuracy at a mean level (i.e., elevation) depends critically on the criterion adopted and that high and low self-esteem individuals do not differ in terms of how well they fare on alternative measures of accuracy such as scatter and shape (Cronbach & Gleser, 1953). Little is known about how sensitive these two types of individuals are to how their own behavior is affected by different social circumstances.

In the present study, we explored the impact of the stimulus manipulation on the behavior and metaperceptions of individuals with positive versus negative feelings about themselves. We expected that although the behavioral responses of these two groups to the stimulus manipulation would be parallel, the effects on their metaperceptions would diverge. Those individuals who felt dissatisfied and who rated themselves as low in self-confidence should be particularly unlikely to recognize how the positivity of their self-presentation was affected by the questionnaire that they read prior to their own interview.

**Method Participants**

One hundred six University of Manitoba students (60 women and 46 men) enrolled in Introductory Psychology received course credit for their participation in the 1-hr. experiment. Participants were scheduled in same-sex pairs. Prior to each session, we ascertained that pair members were unacquainted with one another. In addition, because the student population was quite heterogeneous, we ensured that pair members shared a common ethnic background and that neither pair member had been attending university for more than 4 years. The 53 pairs were distributed across the four cells of the 2 (positive vs. negative stimulus) × 2 (partner vs. nonpartner) design (n = 13, 14, 14, and 12 for the respective cells), with the ratio of women to men held roughly constant across the cells.

**Pretest**
The participants in this study were randomly selected from a larger pool of students who had completed self-ratings as part of a mass testing session held 2 to 3 months prior to the experiment. The self-ratings were made along the same five satisfaction dimensions used in Study 1, except that the "adjusted to university life" item was replaced with "satisfied with my intellectual abilities," and the judgments were made relative to the average University of Manitoba student. These ratings were made on 9-point scales on which higher numbers reflected more positive perceptions of the impressions conveyed.

Procedure

The procedure was identical to that of Study 1. The only difference was that for half of the pairs (those in the nonpartner condition) the actor was told that the stimulus questionnaire belonged to "someone who participated in this study a few weeks ago." Actors in the nonpartner condition otherwise received the same instructions as those in the partner condition.

Dependent Measures

As in Study 1, participants were first asked a filler question about their preference for different response formats. Actors proceeded to complete the metaperception and then the self-rating questions; observers indicated their impressions of the actor. The dimensions were modified to correspond to the five questions administered in the pretest; participants made their ratings on 7-point scales. In addition, the instructions for the metaperception questions were changed to emphasize to actors that they should "indicate the impressions that you think an outside observer (i.e., someone who has never met you) would form of you, on the basis of your tape." Finally, as a manipulation check, actors rated "the person whose experiences you read about" along the same five dimensions. Although actors always completed the metaperception items first, the order in which they filled out the ratings of their true self and these manipulation check items was counterbalanced.

Results

Preliminary analyses revealed no significant main effects for gender nor any interactions between gender and the experimental manipulations. This variable is therefore excluded from the analyses reported below.

Manipulation Check

To examine whether the stimulus manipulation was effective, we entered actors' impressions of the person they read about into a 2 (positive vs. negative stimulus) x 2 (partner vs. nonpartner) ANOVA. This analysis confirmed that actors in the positive condition formed a more positive impression of the person they read about (M = 5.37) than did actors in the negative condition (M = 2.49), F (1, 48) = 503.17, p < .001. There were no other effects.

Replication of Study 1

Our initial analysis of the satisfaction data had two key objectives. First, we sought to assess whether actors' metaperceptions and observers' impressions were differentially affected by the stimulus manipulation; such a pattern would replicate the findings of Study 1. Second, we wanted to explore whether this effect was contingent on the stimulus questionnaire belonging to actors' interaction partner. Participants' responses along the five satisfaction dimensions were combined to form indexes, and these data were analyzed in a 2 (actor vs. observer) x 2 (positive vs. negative stimulus) x 2 (partner vs. nonpartner) repeated measures ANOVA, with pairs as the unit of analysis. The first factor was a within-pairs variable, and the remaining factors were between-pairs variables. Consistent with our hypotheses and the results of Study 1, this analysis yielded a significant Actor—Observer x Stimulus interaction, F (1, 49) = 4.20, p < .05. Simple effects analyses revealed that actors in the positive stimulus condition conveyed more positive impressions (M = 4.99) than did actors in the negative stimulus condition (M = 4.31), F (1, 78) = 7.50, p < .01, but they did not perceive that they did so (M = 4.40 and 4.21 for metaperceptions in the positive and negative conditions, respectively), F (1, 78) < 1. It is also possible to examine the interaction by comparing actors' metaperceptions to observers' impressions in each of the two stimulus conditions. These analyses revealed that whereas in the negative condition there was no systematic actor—observer difference, F (1, 49) < 1, in the positive condition actors' metaperceptions were less positive than observers' impressions, F (1, 49) = 11.12, p < .005. The overall analysis also yielded a main effect for the stimulus manipulation (M = 4.70 and 4.26 for the positive and negative conditions, respectively), F (1, 49) = 4.03, p = .05, and a main effect for the actor—observer variable (M = 4.30 and 4.65 for actors and observers, respectively), F (1, 49) = 7.98, p < .01, each of
which was qualified by the interaction described above. There were no other effects. Most notably, participants’ perceptions were not significantly affected by whether the stimulus questionnaire ostensibly belonged to the observer or to someone outside of the current interaction context. As in Study 1, we computed the effect size for the influence of the stimulus factor on observers’ impressions to examine the impact of the manipulation more precisely. The value obtained after correction for overestimation (.73) again was considerable. Finally, actors' metaperceptions and observers' impressions were significantly correlated, average within-cell $r (41) = .51, p < .001$, even though they were differentially affected by the stimulus manipulation.

**Correlations With Actors' Prior Self-Knowledge**

Next, we examined the relationship of actors' metaperceptions and observers' impressions to actors' prior self-knowledge. We expected that actors' metaperceptions would be clearly connected to their general self-knowledge. Although such a relationship would not demonstrate in any definitive way that self-knowledge hindered actors' ability to detect the influence of the stimulus manipulation on their behavior, it would be consistent with this possibility. Given observers' relatively limited access to actors' private self-knowledge, we expected a weaker relationship between observers' impressions and actors' prior self-ratings. Correlational analyses performed with actors' prior self-ratings (combined across the five dimensions) confirmed these predictions. Although actors' metaperceptions were related to their prior self-ratings, average within-cell $r (40) = .42, p < .01$, observers' impressions were not, average within-cell $r (40) = .17, ns$. These correlations were significantly different, $t (37) = 2.40, p < .05$. In sum, the impressions that actors conveyed to the observer were connected to the stimulus manipulation, whereas actors' perceptions of the impressions that they conveyed were connected to their general feelings of satisfaction with their personal accomplishments and academic experiences. These findings suggest that in estimating the impressions that they conveyed during their interview, actors may have overlooked a powerful influence on their current behavior in favor of self-knowledge that was relatively inaccessible to observers.

**Actors' Self-Ratings**

To examine whether actors felt that there was a discrepancy between their true self and the self that they presented, and whether their sense of discrepancy was affected by the stimulus manipulation, we combined actors' self-ratings along the five satisfaction dimensions and entered these data into a 2 (true self vs. metaperception) × 2 (positive vs. negative stimulus) × 2 (partner vs. nonpartner) repeated measures ANOVA. The first factor was a within-subject variable, and the remaining factors were between-subjects variables. The results were consistent with the findings of Study 1. Actors generally felt that their true self was more positive ($M = 4.54$) than the impression they had conveyed ($M = 4.32$), $F (1, 48) = 7.60, p < .01$. This effect was not qualified by either of the experimental manipulations ($F$’s $< 1$). As was the case in Study 1, a direct examination of actors’ ratings of their true self revealed that actors in the negative and positive stimulus conditions rated themselves quite similarly ($M = 4.45$ and $4.62$, respectively). As would be expected, actors' self-ratings were significantly related to their prior self-ratings from the pretest, average within-cell $r (40) = .52, p < .001$.

**Individual Differences**

In a final set of analyses, we explored whether actors’ sensitivity to changes in their self-presentations was connected to the positivity of their general self-evaluations. Actors' self-ratings from the pretest served as indicators of their general self-esteem. We performed a median split on the self-rating data, such that students who received a mean score that was less than or equal to 5 (which was in fact the midpoint of the 9-point scale) were assigned to the low self-esteem group ($n = 27$), and those who received a score greater than 5 were assigned to the high self-esteem group ($n = 26$). We then included this self-esteem factor in an analysis comparing actors' metaperceptions and observers' impressions: We examined the data with a 2 (actor vs. observer) × 2 (positive vs. negative stimulus) × 2 (high vs. low self-esteem) ANOVA. Pairs was the unit of analysis; the first factor was a within-pairs variable, and the remaining factors were between-pairs variables. We excluded the partner factor from this analysis because we had no hypotheses regarding how it might interact with actors' self-esteem, and there was no indication from our other results that it had been influential. The resultant cell sizes were slightly uneven but acceptable ($n = 11, 16, 16$, and $10$). The analysis yielded an overall Actor—Observer × Stimulus × Self-Esteem interaction, $F (1, 49) = 3.30, p = .076$, suggesting distinct effects for pairs with a high versus a low self-esteem actor. Consistent with our predictions, simple effects analyses revealed that the blindness effect was produced primarily by those actors who were relatively low in self-esteem. For pairs with a low self-esteem actor, there was a significant Actor—Observer × Stimulus interaction, $F (1, 49) = 5.97, p < .025$. Actors who were low in self-esteem conveyed more positive impressions in the positive ($M = 4.96$) as opposed to the negative ($M = 4.09$) stimulus condition, $F (1, 78) = 6.55, p = .01$, yet their metaperceptions did not vary across the two conditions ($M = 4.04$ and 4.00
for the positive and negative conditions, respectively; \( F < 1 \). For pairs with a high self-esteem actor, results indicated a main effect for the stimulus manipulation, \( F(1, 78) = 5.94, p < .025 \), that was not qualified by the actor—observer variable (interaction \( F < 1 \)). Thus, actors who were high in self-esteem conveyed more positive impressions in the positive than in the negative stimulus condition, and they were sensitive to this change ( \( M = 5.01 \) and 4.41 for the positive and negative conditions, respectively; means were combined across actors' metaperceptions and observers' impressions). The cell means are presented in Table 1. Finally, consistent with the idea that low self-esteem individuals' self-concepts exert a particularly powerful influence on their metaperceptions, correlational analyses revealed a stronger relationship between prior self-ratings and metaperceptions for low self-esteem actors, average within-cell \( r(20) = .50, p < .025 \), than for high self-esteem actors, average within-cell \( r(20) = .21, ns \). Although the difference between these correlations was not statistically significant ( \( z = .98, ns \)), the pattern suggests that low self-esteem may be associated with an enhanced influence of personal self-beliefs on metaperceptions during social interaction.

### Discussion

The results of Study 2 replicate and extend the findings of Study 1. Actors matched their own self-presentations to the self-description they read prior to their interview, yet they did not recognize how their behavior had been influenced. It is notable that this effect was obtained with the revised instructions to actors, which emphasized that they should consider an outside observer's reactions to their remarks. An account for actors' insensitivity in terms of corrections for the observer's frame of mind or understanding of his or her own influence was thus rendered untenable.

Actors matched their self-presentation to the self-description in the questionnaire regardless of whether it ostensibly belonged to their observer partner. This suggests that their behavior might reflect adherence to an implicit social norm that the self-description set regarding the appropriate types of experiences to discuss in the experimental setting (e.g., that it was acceptable in this context to complain and share failure experiences or instead to speak positively about one's self). Alternatively, behavior matching may have occurred because the different self-descriptions primed different aspects of actors' own experiences. Regardless, actors' responses to the manipulation did not seem to depend on motivations centering on the feelings or reactions of their observer partner. In our view, this says less about people's concern for others' feelings than it does about the minimal conditions for behavior matching.

Study 2 suggests one source of information that actors may have used when estimating the impressions they conveyed with their self-descriptions: Actors' metaperceptions were closely related to their general self-knowledge (as assessed months prior to the experimental session), even though observers' actual impressions were not. This finding suggests that actors may have relied on their general feelings about their academic accomplishments and university experiences when judging how they appeared. Observers' impressions were instead tied to actors' current behavior, which was guided by the stimulus manipulation.

Finally, the results of Study 2 suggest that individuals high and low in self-esteem are differentially sensitive to how they engage in behavior matching with respect to the positivity of self-presentation. Although individuals high in self-esteem were aware of how their behavior was affected by the stimulus manipulation, individuals low in self-esteem were not. Indeed, participants who were low in self-esteem demonstrated a remarkable failure to discern how the positive stimulus questionnaire prompted them to present themselves in a favorable light. The pattern of means was thus consistent with a self-protective account for the insensitivity demonstrated by low self-esteem participants. At first blush, these results may seem to contradict research indicating greater temporal stability and clarity in the self-concepts of high as opposed to low self-esteem individuals (Campbell, 1990; Campbell et al., 1996; Kernis, Cornell, Sun, Berry, & Harlow, 1993).

However, it is important to recognize that the present studies addressed the separate issue of individuals' sensitivity to when they have been led astray from their self-concept, such that they present themselves in a manner that is contrary to their overall self-evaluation. High self-esteem individuals' greater attentiveness to diagnostic self-relevant information (Campbell, Fairey, & Fehr, 1986) may render them more likely than their avoidance-oriented low self-esteem counterparts to consider the implications of their current behavior.

Our results should be considered tentative in light of the fact that they were not obtained with a traditional measure of self-esteem. In addition, there are other possible interpretations of the association, aside from the self-protective account that we have emphasized. For example, low self-esteem might be a product of how individuals process information about their own behavior or of poor perspective-taking skills. The findings do, however, merit further attention. In particular, they may have implications for self-concept change. It would seem that a negative self-view might serve as a perceptual barrier at a level deeper than attributions, preventing individuals from even identifying behavior with positive implications in the first place. In this regard, it would be informative to explore the generality (or specificity) of low self-esteem...
individuals' insensitivity.

General Discussion

The present research demonstrates that individuals can fail to perceive changes in their actions that are induced by implicit social influence. In both Studies 1 and 2, actors matched the positivity of their self-presentations to the positivity of the self-description that they read immediately prior to their interview but did not perceive that they had done so. The results from the transcript coding conducted in Study 1 suggest that actors' and observers' differential access to information contained in nonverbal channels could not fully account for these findings. The findings of Study 2 suggest that actors' inferences were connected to their (currently) nondiagnostic general self-concept rather than to the implications of their present behavior, on which observers' actual impressions were based. The present results are consistent with recent research demonstrating that people do not appreciate how they are perceived differently by different interaction partners (DePaulo, Kenny, Hoover, Webb, & Oliver, 1987; Kenny & DePaulo, 1993; Schechtman & Kenny, 1994). Our research extends this finding to situations in which individuals are the target of social influence. In addition, an accurate judgment about how one is viewed by a particular other person depends to some extent on attentiveness to the social feedback that is available from the person as well as on an appreciation of his or her unique point of view. In our experiments, we demonstrated that individuals are also insensitive when making judgments restricted to the implications of their own behavior.

Does Self-Knowledge Cause Blindness?

We have characterized actors' insensitivity to how they are affected by implicit social influence as being due to the blinding effect of self-knowledge. However, although we view the process as one of active interference, our data do not definitively demonstrate that self-knowledge causes actors to fail to recognize when they engage in behavior matching. Rather, we have shown a connection between actors' metaperceptions and general self-knowledge "side by side" with their insensitivity to the implications of their current behavior. It is instructive here to consider the ways in which actors are different from observers, as these differences represent the potential causes of the actor—observer discrepancy that has been demonstrated in the present research. Three general categories of differences have typically been identified (Bem, 1972; Jones & Nisbett, 1972): (a) physical perspective, (b) background information (including historical information about past behavior as well as current access to internal experiences such as physiological sensations), and (c) motivation to view the self positively. We considered the implications of physical perspective in Study 1, in which we explored the role played by access to the actor's nonverbal behavior and found that this factor could not fully account for the discrepancy between actors' and observers' judgments. We should note that in attribution research, physical perspective is often construed in terms of perceptual salience: Actors' focus on the situation leads them to explain their behavior in terms of the situation, whereas observers are more inclined to make attributions to the perceptually prominent actor. It is unclear how this analysis could be applied in the present context, in which we explored behavior identification: The situation ought to have provided a clear cue to actors as to the type of behavior that they were likely to exhibit. Indeed, actors were at an advantage in that they had access to the stimulus for their behavior, whereas observers did not. Thus, although actors may look to the situation to explain their actions, the present research suggests that they do not readily look there for cues as to how to characterize the nature of their behavior.

Of the two remaining categories of actor—observer differences, we have focused on background information rather than motivation as an explanation for actors' insensitivity to how they are affected by implicit social influence. Although self-enhancement motivations could account for a failure to perceive that one has conveyed a negative impression, it would be difficult to argue that such motivations could explain a failure to perceive that a positive impression has been communicated. However, there is a different type of motivation that could potentially account for the pattern of actors' inferences, one that is consistent with our analysis in that it accords a pivotal role to the nature of individuals' beliefs about themselves. Swann (1987) has argued that individuals possess a motivation for self-verification: People want others to see them as they see themselves, as this affords them a sense of existential security and the ability to predict and organize their social experiences. This motivation would coincide with the self-protective concerns of low self-esteem individuals: Presumably these individuals are hesitant to perceive that they have conveyed positive impressions because they suspect that such impressions might be false, and they sense costs associated with being perceived more positively than is warranted. If the self-verification motive does account for the effects obtained in the present research, it is interesting that its impact was evident with respect to actors' metaperceptions but not with respect to their actual behavior. In sum, the data from the present research suggest that actors' self-knowledge may impede their ability to detect when they engage in behavior matching. The power of individuals' general beliefs about themselves to "engulf the field" could derive from the extensive and well-articulated nature of this information, from the motivations associated with it, or both.
Future research in which individuals are categorized in terms of whether they possess a self-schema (Markus, 1977) relevant to the interaction would provide a more definitive test of whether self-knowledge interferes with sensitivity to behavior matching.

**Self-Presentation Research**

Our demonstration of actors' insensitivity to how they are affected by implicit social influence would seem inconsistent with the results of investigations exploring the internalization of strategic self-presentation. Numerous researchers have illustrated that when individuals are induced to present themselves in a particular way, their self-concepts shift in the direction of their current self-presentation. Although people's perceptions of their own behavior were not directly examined, it is difficult to imagine that self-concept change could occur if people were insensitive to the implications of their current actions and remarks. How, then, can the present findings be reconciled with this literature?

Experiments demonstrating the internalization of self-presentation have typically incorporated a number of procedures that seem apt to facilitate individuals' recognition of changes in their behavior. In many of these studies, participants received explicit and direct instructions from the experimenter to behave in a particular fashion. For example, individuals have been told specifically how independent they should appear to be (Schlenker & Trudeau, 1990), either to be self-deprecating or self-enhancing in their self-presentations (Jones, Rhodewalt, Berglas, & Skelton, 1981, Experiment 3; Rhodewalt & Agustsdottir, 1986; Upshaw & Yates, 1968), whether they should behave as though they are high in emotional stability or high in emotional responsiveness (Tice, 1992, Experiment 1), to act in either an introverted or extroverted manner (Tice, 1992, Experiments 2 and 3), or that there are concrete rewards (e.g., money or a desired job) for making a positive impression on their audience (Jones et al., 1981, Experiments 1 and 2). Alternatively or in addition, participants' behavioral options have been clearly quantifiable: After a brief open-ended preamble (or none at all), the behavior sample has consisted of participants making numerical ratings of themselves on adjectives and descriptive phrases relevant to the target dimension (Jones et al., 1981, Experiments 1—3; Rhodewalt & Agustsdottir, 1986; Schlenker & Trudeau, 1990; Upshaw & Yates, 1968). These aspects of the research methodology seem quite likely to enhance participants' propensity to recognize induced changes in their behavior.

An investigation by Baumeister, Hutton, and Tice (1989), although outside of the internalization literature, is also relevant here. Baumeister et al. developed a paradigm that involved both implicit social influence and a relatively flexible behavioral repertoire. Protagonists were instructed to engage in self-promotion or self-disparagement when they and a naive partner were questioned by an interviewer. Although the study was not designed to test an awareness-of-being-influenced hypothesis, exploratory dependent measures assessed each person's beliefs about how he or she would be rated by his or her partner. Thus it is possible to look at the naive partners' actual and self-perceived responsiveness to the protagonist's behavior from this perspective. As we would expect from the implicit nature of the influence and the relatively extensive behavioral repertoire available to participants, the naive partners did not appear sensitive to the fact that they had been influenced by the behavior of the protagonist. However, although these results were intriguing, the authors themselves acknowledged that aspects of the procedure rendered them inconclusive. Our investigations provide a more definitive demonstration of individuals' blindness to how they are affected by implicit social influence.

**Attitudes and Expectations: Implications for Pluralistic Ignorance and Self-Fulfilling Prophecies**

Although the present experiments were based on behavior matching with respect to the positivity of self-presentation, our results regarding individuals' insensitivity to how they are affected by implicit social influence may have implications across a wide range of behavioral domains. In particular, a failure to recognize how one's remarks about an issue conform to the attitudes expressed by one's interaction partners might contribute to the phenomenon of pluralistic ignorance. A state of pluralistic ignorance exists when people believe that others' attitudes and feelings are different from their own, even though their own and the others' public behaviors are similar (Miller & McFarland, 1987, 1991; Prentice & Miller, 1993). Theoretical analyses of pluralistic ignorance have identified two key possible accounts for the phenomenon, differential interpretation and differential encoding (Miller & McFarland, 1991). According to the differential interpretation hypothesis, the phenomenon may occur because individuals make different attributions for their own versus others' adherence to group norms, as a function of believing that they themselves are uniquely vulnerable to fears of being rejected by their peers. According to the differential encoding hypothesis, pluralistic ignorance may arise because people fail to realize the extent to which they appear to support prevailing social practices and viewpoints. Although indirect evidence for the differential interpretation hypothesis has been obtained in experimental research (Miller & McFarland, 1987), the differential encoding (or "illusion of transparency") hypothesis has not been examined. By demonstrating that individuals can be insensitive to how they engage in behavior matching, the present studies provide evidence supportive...
of the differential encoding hypothesis. The implications are clear. If people overlook a schism that exists between what they really feel or think and how they appear, it is unlikely that they will appreciate the possibility of this type of private—public discrepancy in others. Realizing the extent of one's own tendency to engage in behavior matching would seem a first step toward understanding the range of attitudes and opinions that might underlie other people's public behaviors.

Consider in addition the implications of the present findings for the self-fulfilling prophecy phenomenon. To date, research on self-fulfilling prophecies has not directly examined individuals' awareness of how their actions and remarks are affected by their interaction partner's behavior toward them. In cases in which self-concept change occurs, it seems clear that people see impressions that are incongruent with their (initial) self-views. In circumstances in which self-concept change is not evident, however, individuals' metaperceptions are more difficult to deduce. Do women induced to be unfriendly and aloof by an interaction partner who believes them to be physically unattractive recognize their coldness but attribute it to the other person's lack of warmth? Or are they unaware of how they appear? Similar questions arise with respect to individuals whose behavior is affected by the racial or gender stereotypes of their interaction partner. Our understanding of self-fulfilling prophecies might benefit from a consideration of targets' perceptions of their own behavior, in addition to their attributions for it. In particular, the likelihood of any form of reactance against influence would seem to depend first and foremost on targets' recognition of the (potential) impact and then on other factors that have been identified, such as the costs associated with disconfirming the other person's expectations.

Conclusion

Undoubtedly, cues are sometimes available in real-world interactions that can serve to facilitate actors' detection of changes in their behavior that are induced by the situation. For example, explicit feedback from an interaction partner or theories about situations (e.g., job interviews) or relationships (e.g., close friendships) in which accommodation is likely may sensitize people to how they modify their actions and remarks, leading to a greater awareness of behavior matching. Our goal in designing the paradigm for our research was to create conditions that closely mirrored the many everyday circumstances in which such cues to influence are not available, that is, situations in which social influence is of a more implicit nature. We found that under these conditions actors can be unaware of how they engage in behavior matching. These results have clear implications for research exploring self-concept malleability. Moreover, in our opinion, the failure to consider the sensitivity of self-perceptions to behavior change has been a significant oversight in previous investigations of social influence. Behavior matching may sometimes have greater consequences when individuals fail to detect it than when they do.

APPENDIX A

A Negative and Positive Stimulus Self-Descriptions

Negative Stimulus Positive Academic Experiences

I guess I don't really feel like I have had very many positive academic experiences since I came here. Actually, I can't think of anything specific to say for this part. It's been much harder for me than high school was. I've found a lot of the material very difficult, and my grades are low (especially compared to the other people I know here).

Negative Academic Experiences

The worst moment I can think of was my French final; I went completely blank at the start, even though I had studied quite a lot for it. I ended up doing really badly—I got a C in the course. I try not to let this kind of thing get to me too much, but sometimes you can't help it. My parents put a lot of emphasis on grades.

The couple of times when I've spoken up in class, I didn't feel that my ideas went over very well. I tend to take things really personally. People usually seem to ignore or criticize what I say. It's hard when it takes a lot for you to work up to saying something, and then you get the sense that nobody is listening.

—pretty strong negative impact on how I see myself.

Positive Social Experiences

I haven't made many new friends since I got to Princeton. Mostly, I have to rely on the people that I knew before (with phone calls, letters, etc.), but it's hard to keep these relationships going long-distance. It is nicer to have people "right there" when you need them. But I just don't feel as though I have much in common with students here. I don't feel like
they really accept me.

Negative Social Experiences

A couple of times I have tried to take the initiative with people (by suggesting that we do things together), but I often end up feeling like they are trying to put me off, or avoid me. They don't seem very enthusiastic about getting to know me. When I do go out, I tend to feel awkward, especially in groups. I get stuck without anyone to talk to, or end up not being able to think of the right thing to say. There are a couple of people that I can talk to, but I can tell that I am not really a "high priority" for them.

It doesn't help that my boyfriend broke up with me just before I came here. I guess it really didn't make sense for us to keep seeing each other, with the distance and everything, but I miss him a lot. If I knew back then how I would feel now, I think I would have tried to convince him that we should keep things going on some level.

So I'm having trouble dealing with the loneliness.

Positive Stimulus Positive Academic Experiences

I feel pretty good about most of the academic experiences I have had here. It is definitely quite different from high school—not necessarily in terms of the level of intensity, since I went to a pretty demanding high school—but more in terms of the increased amount of time spent working independently, as opposed to class time. I really like this change. I'm doing well in my courses, and I'm getting pretty good grades on exams and papers.

My experience at Princeton has, I think, greatly enhanced my ability to think critically. Through precepts I have learned to verbalize my thoughts, and have in general become much more comfortable and interested in discussing things on an intellectual level.

Overall, my academic experiences here have bolstered my self-esteem.

Negative Academic Experiences

I still struggle a bit with speaking in class, though as I said before, it is getting a lot easier. It depends a lot on the size of the group (large groups are the most difficult).

I wish I were criticized a bit more. In particular, I think my writing has suffered from a lack of constructive criticism.

I find I sometimes have difficulty applying myself (i.e., lack of self-discipline).

Positive Social Experiences

I have had some wonderful friendships and roommates. I have really enjoyed just talking to the people I have met here—there are a lot of incredibly motivated people on campus, and it can be contagious. Many of the students have had interesting experiences. This makes for good conversations and creates, in me, a desire to broaden my own horizons as much as I possibly can.

I have found that I generally have more in common with people here than I did with the people I knew in high school. I like the fact that everyone here puts a great deal of effort into something (e.g., academics, athletics, an instrument). I feel more socially accepted than I used to. Probably because I seem to be able to be more outgoing in this atmosphere.

Negative Social Experiences

I have had some soured relationships, which have made me more cautious and a bit unsure of myself re: dating. I have realized that sometimes I am not as sensitive to other people as I could be.

There is a lot of social pressure on campus. I actually think it is far worse than the academic pressure.
I do miss my family. At first I was so caught up in new things here that I didn't notice it as much. I feel it a bit more now, so I am calling them more often.

References


Campbell, J. D., Fairey, P. J. & Fehr, B. (1986). Better than me or better than thee? (Reactions to intrapersonal and interpersonal performance feedback. *Journal of Personality, 54*, 481–493.)


Personality and Social Psychology, 49, 1199—1206.)
Sakurai, M. M. (1975). Small group cohesiveness and detrimental conformity. (Sociometry, 38, 340—357.)
The dependent measures were designed to capture two components of participants' perceptions corresponding to the two sections of the interview, an academic component (Items 1 and 4) and a social component (Items 2, 3, and 5). The patterns of means for these two components were very similar when they were considered separately, and including the components as a within factor in analyses served only to slightly strengthen the main findings (it did not qualify any of the results). This was true for parallel analyses of Study 2 as well. For the sake of simplicity, we therefore combined all of the items together into a single index. In Study 1, the average reliability across the separate academic and social components was .70 for metaperceptions and .62 for impressions. The reliability of the combined index for actors' metaperceptions was moderate (α = .48), as it was for observers' impressions (α = .59). The lower reliability of the combined index reflects an unexpected lack of relation between the academic and social components for both metaperceptions, r(35) = −.16, ns, and impressions, r(35) = .10, ns, in this student population; research on other student samples has sometimes found independence between academic and social components of self-esteem (e.g., Fleming & Watts, 1980). In Study 2, in which we used a different student population (and replaced one of the items), the academic and social components were significantly related for both metaperceptions, r(53) = .39, p = .004, and impressions, r(53) = .37, p = .006, and the reliabilities of the combined indexes were considerably higher. Factor analyses conducted separately for metaperceptions and impressions within Study 1 and Study 2 (involving a total of 176 participants) consistently yielded two factors corresponding to our expectations about the items that would reflect the academic and social components of participants' self-presentations. Consistency over time in actors' and observers' self-ratings (Study 2) and correlations between actors' and observers' perceptions (Studies 1 and 2) further attest to the meaningfulness of these judgments to participants.

The degrees of freedom and error terms used to test simple effects in analyses incorporating between-subjects and within-subjects variables are calculated according to Howell (1987), who recommended a procedure based on Satterthwaite (1946). The procedure leads to unusual values for degrees of freedom when a simple effect partitions a main effect and an interaction effect that were themselves tested by different error terms (which is the case for our simple effects tests of the between-pairs stimulus factor).

The degrees of freedom for all averaged correlations are calculated according to Mather's (1960) recommendations.

The patterns of means for academic and social experiences were virtually identical, and the results were unchanged when these were entered as separate components in the analysis.

Arguably, the audio components of the videotape (i.e., vocal cues such as speed of speech and intonation together with content) most closely approximate the information available to actors themselves. When we had (new) judges make ratings from audiotapes, the results were the same as for the transcripts: There was a significant effect for the stimulus manipulation on impressions of satisfaction but not on percentage ratings.
One participant failed to complete this section of the questionnaire.

The reliability of the combined index of actors' metaperceptions was solid ($\alpha = .79$), as it was for observers' impressions ($\alpha = .79$). The average reliability across the separate academic (Items 1, 4, and 5) and social components (Items 2 and 3) was .72 for actors' metaperceptions and .78 for observers' impressions.

We also explored whether observers' self-ratings from the pretest predicted the impressions that they formed of the actor; there was no significant relationship, $r (53) = .13$, ns. It is notable that results from a supplementary dependent measure that asked observers to describe their true self along the same dimensions as actors revealed considerable temporal stability in these feelings: Observers' pretest self-ratings were significantly correlated with the self-ratings they made at the time of the experimental session, average within-cell $r (41) = .64$, $p < .001$.

Note that a social comparison (or contrast) effect would be problematic for our interpretation of the findings, as it would pose the possibility that actors were sensitive to changes in their behavior but that their sensitivity was masked by other forces operating on their perceptions. The results from Studies 1 and 2 for self-ratings suggest that actors were not affected by social comparison: When actors judged their true self rather than their current behavior, any opposing influence of current behavior ought to have been removed such that social comparison effects were clearer (they were not). To further explore the possibility of contrast effects, we conducted a version of the study in which participants received the stimulus information but did not engage in any self-presentation. Sixty-four participants read a scenario in which they had a conversation with another student who had been at university the same number of years that they had; the person conveyed to them all of the information contained in either the positive or negative questionnaire. Participants formed greatly different impressions of the person in each case ($M = 2.17$ and 4.76 for the positive and negative conditions, respectively), $t (62) = 13.32$, $p < .001$. Their current self-ratings were unaffected by which version they read ($M = 4.36$ and 4.15 for the positive and negative conditions, respectively), $t (61) < 1$. Further evidence against contrast effects is provided by the average within-cell correlations between actors' metaperceptions and impressions of the stimulus questionnaire. Controlling for observers' impressions (a proxy for actors' actual behavior), the average correlations for Studies 1 and 2 were $r (29) = -.14$, ns, and $r (40) = .53$, $p < .001$, respectively; the results were unchanged when simple correlations were calculated instead. The positive relation in Study 2 is clearly inconsistent with the possibility that actors' metaperceptions were guided by contrasts with the stimulus questionnaire. The data from observers can also be used to address the possibility of contrast effects. Across both studies, the only significant associations between observers' current self-ratings and impressions of the actor were positive rather than negative. Taken together, these analyses make it difficult to maintain that a contrast mechanism was behind the effects obtained across the two studies. Note, finally, that the results for the percentage rating measure are not vulnerable to a contrast interpretation, as responses on this measure constituted absolute characterizations of behavior; moreover, the manipulation check demonstrated that participants in both conditions saw the stimulus person as atypical (i.e., not representative of the norm).

The average within-cell correlations between actors' prior self-ratings and judgments of their true self were $r (19) = .44$, $p < .10$, and $r (20) = .29$, ns, for low and high self-esteem individuals, respectively ($z = .52$, ns).
Fazio, Effrein, and Falender's (1981) demonstration that self-fulfilling prophecies can ultimately affect targets' self-concepts also involved a manipulation that was directive in nature. Participants were subjected either to a series of eight questions designed to elicit feelings of extroversion (e.g., "What would you do if you wanted to liven things up at a party?") or to a series of eight questions designed to elicit feelings of introversion (e.g., "What things do you dislike about loud parties?"). Participants' subsequent self-ratings and behavior were influenced by this manipulation.

Table 1.