Journal of Personality and Social Psychology
1996, Vol. 71, No. 4, 703-716

0022-3514/96/$3.00

Truth and Investment: Lies Are Told to Those Who Care

Bella M. DePaulo
University of Virginia

Kathy L. Bell
Texas Tech University

Participants discussed paintings they liked and disliked with artists who were or were not personally invested in them. Participants were urged to be honest or polite or were given no special instructions. There were no conditions under which the artists received totally honest feedback about the paintings they cared about. As predicted by the defensibility postulate, participants stonewalled, amassed misleading evidence, and conveyed positive evaluations by implication. They also told some outright lies. But the participants also communicated clearly their relative degrees of liking for the different special paintings. The results provide new answers to the question of why beliefs about other people's appraisals do not always correspond well with their actual appraisals.

In their formal roles as parents and supervisors, and in their informal roles as colleagues and friends, people often provide us with evaluative feedback. They comment on our work, our behavior, our friends, and our lovers. These appraisals are important for many reasons, including three interdependent ones. First, evaluative feedback can be of great emotional significance. Second, it can have instrumental value; for example, it can shape performance and guide important life decisions. Third, the appraisals of others—or our perceptions of them—can form and inform our sense of self (e.g., Baldwin, 1992; Felson, 1992; Jussim, Soffin, Brown, Ley, & Kohlhepp, 1992; McNulty & Swann, 1994; Mead, 1934). According to the symbolic interactionists, the self that develops is a "looking glass self" (Cooley, 1902) formed by our perceptions of others' responses to us.

The looking glass metaphor seems to imply that the question of the accuracy of perceptions is nonproblematic; we can simply look to others and see their opinion of us reflected back to us directly (Felson, 1992; Glaser & Strauss, 1964). Yet the preponderance of evidence suggests that there is considerable error in our perceptions of how others view us (e.g., DePaulo, Kenny, Hoover, Webb, & Oliver, 1987; Kenny & DePaulo, 1993). Our perceptions of others' appraisals correspond imperfectly with their actual appraisals, and our self-perceptions are more closely linked to our perceptions of how others view us than to their actual views of us (Felson, 1992).

When we err in our perceptions of how others view us, we may do so because others did not communicate their views of us openly and honestly, or because we misinterpreted their appraisals. Three elements are important: what the evaluators tried to convey (as indicated by their own reports), what they actually did convey (as indicated, for example, by transcripts of what they said), and how their communications were perceived (as indicated by people's impressions). Studies of meta-accuracy typically omit the middle element: There is no precise record of what evaluators actually said, or the record is never analyzed. In the present research we assessed all three components.

An important reason for dishonesty in evaluative communications may be that evaluators care more about the emotional impact of their feedback than its instrumental value. The feedback that supervisors can provide to floundering employees, for example, is potentially of instrumental value both to the employees and to the organization, yet supervisors are reluctant to provide feedback to those employees and often delay doing so (Larson, 1989). Honesty and openness are highly prized characteristics of friendships, yet even friends are reluctant to share their unflattering appraisals of each other (Blumberg, 1972; Mayer, 1957). The persons directly affected by bad news have the greatest need to know that news, yet people are more inclined to communicate such news to uninvolved third parties than to the targets (Felson, 1992; Tesser & Rosen, 1975).

As an individual's personal investment in an object increases, both the instrumental and the emotional significance of evaluative feedback are likely to increase as well. For example, when an art student is discussing paintings with other people, the appraisals that they can provide are more emotionally impactful and also more useful when the paintings are the art student's own work than when they are the creations of other artists. Yet we think that emotional considerations will prevail, and evaluators will be less honest about the paintings when they are the art student's own work—particularly when they dislike the work—even though it would be especially useful to the art stu-

Bella M. DePaulo, Department of Psychology, University of Virginia; Kathy L. Bell, Department of Psychology, Texas Tech University.

This investigation was supported in part by a Research Scientist Development Award from the Alcohol, Drug Abuse, and Mental Health Administration and the National Institute of Mental Health (NIMH) and an R01 Award from NIMH. We thank Malaika Almarode, Dan Alter, Lori Cuatina, Robert Cooley, Alexandra Dahne, Irene Dalton, Laura Daniel, Naomi Driesen, Dave Gawrylowicz, Darby Gibbs, Salina Guliani, Joan Hairfield, Joe Ho, Ken Hodge, Alicia Hughes, Melinda Lantz, Traci Mann, Joanne Moak, Nikki Picerno, Lynne Robinson, Sean Robinson, Laura Rogers, Karen Swanner, Laura Tuck, Benita Watson, Ken Watson, and Cheryl Witt for their help with this research.

Correspondence concerning this article should be addressed to Bella M. DePaulo, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22903. Electronic mail may be sent via the Internet to bmd@virginia.edu.

dents to know how their work really is perceived by other people. For instance, art students whose work is poor yet who never hear that from others may pursue a major or even a career to which they are ill suited.

In the present research, participants looked over a set of paintings, chose the two they liked the best and the two they liked the least, and indicated just how much they liked each of those four paintings. They also wrote out what they liked and disliked about each painting. Only then did they learn that they would be discussing those paintings with an art student who was personally invested in one of the liked and one of the disliked paintings. Those conversations were videotaped (and later transcribed). After each conversation, we asked the participants how honest and how comfortable they had been and how much liking they had tried to convey. We then showed the videotapes to judges who indicated their perceptions of the participants' honesty and actual liking for the paintings. We predicted that the participants would be more dishonest and more uncomfortable, and would exaggerate their liking more, when they were discussing the paintings that were special to the art student—especially when they disliked those paintings.

Goffman (1967, 1971) provided a perspective for understanding people's reluctance to say exactly what they feel. He argued that in order for everyday social life to proceed smoothly, it is important for people to give deference to the "faces" (identities) that others seem to be claiming. As politeness theory has documented (P. Brown & Levinson, 1987; R. Brown & Gilman, 1989; Holtgraves, 1992), people understand this and act accordingly. No one needs to tell us to be polite when discussing an ugly painting with the artist who created it. Disagreements and criticisms are face-threatening and will be communicated only very politely, if at all—but even positive communications, P. Brown and Levinson argued, can be face-threatening (e.g., compliments that cause embarrassment).

Telling people explicitly to be polite and to try to avoid hurting another person's feelings, then, should result in communications that are no different than if no instructions had been given. In both instances, people will dishonestly convey overly positive appraisals. To break down the sturdy barriers to the communication of negative evaluations, it may be important to underscore explicitly the importance of honesty. In the present research, we explicitly instructed some of our participants to be honest about their appraisals. Only from such honest evaluations, we said, could the art students really learn about other people's perceptions of art. We predicted that these instructions would dampen participants' exaggerations, relative to conditions in which participants were instructed to be polite or were given no special instructions, but we were unsure as to whether they would elicit evaluations that were totally honest.

The situation we created was a very difficult one for the participants, especially when they were discussing paintings they disliked with the art student who painted them. Bavelas and her colleagues (Bavelas, Black, Chovil, & Mullett, 1990) characterized this situation as the most common sort of "communicative avoidance–avoidance conflict: [Participants had] a choice between saying something false but kind and something true but hurtful" (p. 58). On the basis of more than a dozen experiments, Bavelas et al. concluded that, in these situations, people equivocate. They avoid answering the question that is asked, they avoid describing their own opinion, they are unclear in the answers they do give, and they sometimes even avoid addressing the person who posed the question. Bavelas et al.'s research, then, tells us what people do *not* say in avoid–avoid situations (or at least in role-play versions of them), but it stops short of telling us what they *do* say. Even their conclusions about what people do not say are based not on content analyses of the communications but on judges' global impressions.

We agree with Bavelas et al. (1990) that people prefer to avoid telling either outright lies or hurtful truths. Therefore, we predict, as they did, that the rate of telling outright lies will be low. However, we think that the rate of lying, though low, will still be responsive to our experimental manipulations. Specifically, we predict that participants will be most likely to lie when discussing paintings they dislike with art students who are personally invested in them—especially if the participants had been instructed to be polite.

In the difficult situation we created, we think that participants have two goals: They want to mislead the art student about how they feel, but they also want to be able to deny that they lied. Their communications will be governed by what we will call the *defensibility postulate*, that is, participants' inclination to exaggerate their liking for the paintings and to convey dishonest appraisals of them will be tempered by considerations of defensibility (see also Schlenker, 1980). Participants will craft communications which, if challenged, can be defended as either truthful or at least not clearly deceptive. In the context of this experiment, we think that one way they can do this is to amass misleading evidence. As the art student continues to probe them about their opinions of the paintings, they can mention more and more of the things that they really do like about the paintings, while being a bit more restrained in enumerating the aspects of the paintings that they really do dislike. The result is a communication that is likely to succeed in conveying a misleadingly positive impression yet can still be defended as truthful—after all, all of the positive aspects mentioned were ones that the participants really did like about the paintings. Not mentioning all of the disliked aspects, they might argue, is not dishonest—they just did not mention them.

We think that the participants will also come up with entirely new aspects of the paintings that they will claim to like—aspects that they had not written down when we first asked them to describe what they liked and disliked about the paintings. Perhaps they will tell themselves that they just noticed these new virtues of the painting during the conversation with the art student. Defensibility is especially likely to remain intact if they also notice some new aspects of the painting that they dislike. Again, though, the newly discovered disliked aspects will be far outnumbered by the new liked aspects.

The prediction made by Bavelas et al. (1990) that people will avoid stating their own opinion is consistent with the defensibility postulate and was directly tested by the coding of participants' explicit expressions of liking or disliking for the paintings. When participants are discussing a painting they dislike, especially one that is special to the art student, they might stonewall—that is, avoid making any explicit evaluations at all. They might also mention fewer aspects of the paintings that they like or dislike.

There is another very clever way that participants can defensibly imply more liking than they really do feel for the paintings, and that is by manipulating what they say about the paintings in which the art students are *not* personally invested. That is, at the same time that participants try to avoid saying explicitly that they dislike the art student's own paintings that they detest, they can be far less reticent in voicing their distaste for the paintings created by other art students. The strategy is one of social comparison by implication. In comparison to the negative appraisals that were explicitly stated about the other artists' work, the withholding of any explicit appraisals of the art student's own work will seem rather positive. Those communications are also defensibly positive: If pressed, the participants can claim that they did *not* say that they liked the art student's own work; they simply said that they did not like the other artists' work.

When we showed the videotapes of the conversations to the judges, we gave them the same information that the artists would be likely to have in the comparable real life situations. That is, the judges knew whether the paintings were special to the artists, but they did not know what the participants really did think of the paintings. They also did not know the participants' intentions—that is, whether they were making any special effort to be honest or polite.

We predicted that the judges would report some of the same things that the participants would say themselves—that the participants were less honest and less comfortable, and exaggerated more, when discussing the paintings in which the artists were more invested (cf. DePaulo & Kirkendol, 1989; DePaulo, Lanier, & Davis, 1983; DePaulo, Stone, & Lassiter, 1985b). If that were all that the judges noticed, then those results might simply reflect judges' theories about how people communicate to people who care, rather than any real discernment. However, because the judges did not know whether the participants liked or disliked a painting, if they also thought that the participants seemed especially less honest when discussing the special paintings when the participants disliked those paintings, then they would be showing some insight into participants' true feelings.

It is important to note that we asked the judges directly just how much they thought the participants really did like each of the paintings. If they discounted the participants' expressions of liking too much in the special conditions (because they knew that the participants were talking to artists who were personally invested in the paintings), they would be wrong about the participants' actual feelings (cf. Gilbert & Malone, 1995; Snyder & Frankel, 1976). If instead they were too inclined to take what the participants said at face value (e.g., DePaulo, 1992, 1994; DePaulo, Stone, & Lassiter, 1985a; Gilbert & Malone, 1995; Jones, 1990), they would again be wrong, but in a different direction. That is what we predicted. Because we expected the participants' verbal strategies to be effective in creating misleadingly positive impressions, we expected the judges to believe that the participants really did like the special paintings more than the not-special ones.

Method

Participants and Art Students

Participants were 47 male and 47 female introductory psychology students who participated for partial fulfillment of a course requirement in an experiment that was ostensibly about psychology and art. Five other participants were excluded: 2 men and 1 woman who surmised the purpose of the experiment, 1 man whose speech could not be understood, and 1 woman who completed the forms improperly. Participants were randomly assigned to the six between-subjects cells formed by the crossing of the two degrees of investment (paintings were described as the art student's favorites or her own) with the three kinds of instructions (honest, no instructions, polite). There were 7 or 8 men and 7 or 8 women in each of the cells.

Three women alternated in the role of the art student, and 3 women and 2 men served as experimenters. Preliminary analyses in which art students and experimenters were included as a factor in the design showed fewer significant effects involving the factor than would be expected by chance.

Procedure

Participants were run individually and were told that the experiment was designed to help art students learn more about how art is perceived by people who are not experts. Participants were then left alone in a room to choose the 2 paintings they liked the most and the 2 they liked the least from 19 paintings that were displayed. (The paintings had been painted by undergraduates in an introductory painting course.) Participants rated each of these 4 paintings on 9-point scales of liking, with higher numbers indicating greater liking. The experimenter then returned and gave the participant a second questionnaire on which the participant was asked to describe briefly, in an open-ended format, what he or she liked and disliked about each of the 4 paintings.

The experimenter then told the participant that he or she would now discuss the four paintings with the art student. The experimenter mentioned that the art student may have actually painted some of the paintings herself, and she would tell the participant if she had. The experimenter also informed the participant that the art student would not know that the four paintings were ones that the participant selected and that she would not ever see the participant's ratings of liking for the paintings or the brief descriptions of what the participant liked and disliked about the paintings.

The art student always claimed that one of the participant's two most favorite paintings (randomly selected) and one of the participant's least favorite paintings (also randomly selected) were special to her in some way. The two types of specialness, or *degrees of investment,* were randomly assigned. In the *moderate investment* condition, the art student claimed that the painting was *one of her favorites* ("This is one of my favorites"); in the *high investment* condition, she claimed that the painting was *one of her own* ("This is one that I did."). She introduced this information just before asking the participant what he or she thought of the painting.

Participants were randomly assigned to one of three instructional conditions. One third of them were instructed to be *honest* when discussing the paintings with the art student. Specifically, they were told:

> If it turns out that the art student did paint some of these paintings, you should still be very honest in describing your own opinions about those paintings. Tell her truthfully what you liked and what you disliked about each painting you discuss, even if the paintings are ones she painted herself. This is supposed to be a learning experience for the students. For them to really learn about people's perceptions of art, they have to hear unbiased descriptions of those perceptions. They need to know what you really did like and really did dislike about each painting you discuss.
>
> Some of the art students like to mention which ones they really liked of the ones that are *NOT* theirs. Again, be sure to be honest about your own opinions of the paintings. Tell her what you really

think of the painting, regardless of what her opinions might be. They will learn more if they hear your true opinions.

Another third of the participants were instructed to be *polite* to the art student and to try not to hurt her feelings. Specifically, they were told:

> If it turns out that the student did paint some of these paintings, try to convince her that you really did like the ones she painted so that her feelings won't be hurt. It is OK to mention things you dislike about her paintings when she asks, but just try to convey the impression that overall, you like the ones she did. This study is supposed to be a learning experience for the participants, but we don't want any of them to end up feeling badly because of it.
>
> Some of the art students like to mention which ones they liked of the ones that are *NOT* theirs. Again, it is OK if you don't agree with her—you can say that, but just try to be real nice about it.

In the *no-instructions* condition, participants were not given any particular instructions about what to do.

After determining that the participant understood the instructions, the experimenter left the room, turned on a hidden video recorder, then returned with the art student. After introducing the participant to the art student, the experimenter left the room.

The art student, who was unaware of the participant's instructional condition (but did know which paintings the participants liked and disliked), then proceeded to interview the participant about each of the four paintings, in counterbalanced order. She asked the following questions about each painting, giving the participant ample time to answer each question before moving on to the next: "What do you think of it? What are some of the specific things you like about it? (Anything else?) What are some of the specific things you dislike about it? (Anything else?)" Participants were instructed by the experimenter not to ask the art student about her opinions. The art students were trained to deflect any such questions.

After the discussion of each painting, the art student left the room while the participant completed a questionnaire about the discussion. On 9-point scales, participants indicated how much liking they tried to convey to the art student, how honest and straightforward they had been, and how comfortable they felt while discussing what they liked and disliked about the painting. Participants were debriefed, and all of them signed a consent form allowing us to use their videotapes.

Judges and Videotapes

Seven male and 14 female undergraduates were recruited to rate videotapes (with sound) of the discussions of the paintings. The tapes were rated by just one judge at a time. Not all of the judges rated all of the tapes; on the average, each tape was rated by 6 men (the range was 5–7) and 12 women (11–14).

The discussions of the paintings were edited onto 17 videotapes of about 1 hr each. A nearly equal number of participants from each condition appeared on each tape. After each discussion of each painting, there was a 10-s rating pause. During the pause, judges rated the participant on 9-point scales of honesty, actual liking for the painting, degree of liking that the participant was trying to convey to the art student, and comfort, with higher numbers indicating more of each attribute. Reliabilities (alphas) were .78 for honesty, .94 for actual and conveyed liking, and .71 for comfort.

Transcripts and Coders

Exact typed transcripts were made of all of the discussions of all of the paintings. Three undergraduates coded the transcripts. One coded the conversations of half of the participants in each condition (188 conversations), and a second coded the other half. We used these codings in the analyses. The third person coded 48 of the conversations coded by the first person and 44 of the conversations coded by the second person. The conversations in each set included approximately equal numbers from each condition. The codings of the third person were compared with those of the first two to assess reliability.

Verbal Strategies

Each coder coded three verbal strategies separately for each discussion of each painting.

1. *Explicit evaluations of liking and disliking.* Coders indicated whether the participants explicitly said that they liked the paintings and whether they explicitly said that they disliked them.

2. *Total number of liked and disliked aspects that were mentioned.* Coders counted the total number of different aspects of the paintings that participants said that they liked and the total number they said they disliked. For example, if participants said they liked the color, the shading, and the originality, they would get a score of 3 for total number of liked aspects.

3. *Number of new liked and disliked aspects that were mentioned.* Coders counted the number of aspects of the paintings that participants said they liked and disliked that were different from the aspects that the participants had described in writing before they knew they would be meeting an art student. To code this variable, coders first identified each of the liked and disliked aspects that participants described in writing, then they identified the liked and disliked aspects from the transcripts of the discussions, then they compared the two sets.

Reliabilities

There were two intraclass correlations for each variable: One was the correlation between the first coder and the third, and the other was between the second coder and the third. For explicit evaluations of liking the reliabilities were .96 and .96; for explicit evaluations of disliking they were .85 and 1.00; for total number of liked aspects, .88 and .92; for total number of disliked aspects, .88 and .80; for new liked aspects, .75 and .80; and for new disliked aspects, .71 and .59.

Results

Manipulation Checks

On a manipulation check questionnaire, all participants in the honest condition indicated that their goal was to be honest about their feelings about the paintings. All participants in the polite condition indicated that their goal was to try to be nice and avoid hurting the art student's feelings. When questioned about their understanding of the instructions just before beginning the discussion of the paintings, all participants correctly reported that the art student may have painted some of the paintings (or that some were the art student's favorites) and that they would not know whether the art student had painted any of the paintings until they met her. Ninety-two of the 94 participants correctly indicated that they would be discussing the four paintings they had selected. (Because the 2 participants who initially volunteered the wrong answer corrected it after further probing, and because all of their other manipulation check data were correct, their data were retained in the analyses.) All of the participants understood that the art students would believe that the paintings picked for discussion were selected at random by

Table 1
Effects of Participants' Liking for the Paintings and Artists' Investment on Participants' and Judges' Ratings

	Type of painting								
	Disliked			Liked			*Fs* (1, 82)		
Ratings	Not special	Special	Difference	Not special	Special	Difference	Artists' investment	Participants' liking	Interaction
Participants									
Comfort	6.47	5.05	1.42****	6.88	6.61	0.27	60.19****	58.69****	19.52****
Honesty	7.66	6.38	1.28****	7.99	7.98	0.01	48.50****	67.68****	47.07****
Actual liking	2.30	2.34	0.04	7.35	7.35	0.00	0.03	1392.96****	0.04
Conveyed liking	3.18	3.96	0.78**	6.44	6.79	0.35	26.89****	176.90****	2.90*
Exaggeration	0.88	1.62	0.74***	−0.91	−0.56	0.35	19.94****	74.85****	1.64
Judges									
Comfort	5.85	5.75	0.10**	6.08	6.06	0.02	4.62**	73.70****	3.00*
Honesty	6.56	6.22	0.34****	6.51	6.37	0.14**	46.94****	1.17	6.65***
Actual liking	3.67	3.84	0.17*	5.91	5.98	0.07	4.85**	766.96****	1.11
Conveyed liking	4.14	4.65	0.51****	6.41	6.70	0.29***	45.19****	670.52****	4.24**
Exaggeration	0.47	0.81	0.34****	0.50	0.72	0.22****	84.72****	1.02	3.40*

Note. The special paintings were the ones in which the artists were invested. *Difference* is *not special* minus *special* for comfort and honesty, and *special* minus *not special* for actual liking, conveyed liking, and exaggeration. *MSEs* for artists' investment were, from top to bottom, 2.22, 1.60, 0.82, 2.27, 1.42, 0.26, 0.23, 0.60, 0.66, and 0.08. For participants' liking they were 3.11, 2.59, 1.70, 9.82, 4.93, 0.37, 0.41, 1.18, 1.30, and 0.12. For the interaction they were 3.14, 1.59, 0.77, 2.82, 2.07, 0.24, 0.28, 0.47, 0.54, and 0.09.
* $p \leq .10$. ** $p \leq .05$. *** $p \leq .01$. **** $p \leq .001$.

the experimenter. All participants also understood that the art student would not see what they had written about the paintings.

Analyses of participants' initial ratings of their liking for the paintings indicated that they liked their two favorite paintings far more than their two least favorite ones (*M*s = 7.35 and 2.32). Because the paintings that were described as special to the art student were randomly assigned, participants should not have liked them any better than the ones that were not special, and in fact they did not (*M*s = 4.82 and 4.84; see Table 1 for significance tests).

Design and Measures

Data were analyzed with a mixed-design analysis of variance (ANOVA). The between-subjects factors were the instructional manipulation (participants were told to be honest or polite or they were given no instructions) and artists' degree of investment in the special paintings (those paintings were described as either the artists' favorites—the moderate investment condition, or as their own work—the high investment condition). The within-subjects factors were the artists' investment in the paintings (they were invested in the special paintings and not invested in the not-special ones) and the participants' liking for the paintings (disliking or liking).[1]

Participants' reports of how honest and straightforward they had been in their discussions were highly correlated, $r(92) = .86$, $p < .001$, and so they were averaged to form a single measure of *honesty*. The measure of participants' *actual liking* for the paintings was their ratings of their liking for each of the paintings before they knew that they would be meeting an art student. Their reports of how much liking they had *tried to convey* to the art student were collected after the discussions. We assessed the degree to which the participants had *exaggerated their liking* by subtracting participants' actual liking for each painting from the degree of liking that they tried to convey. The other dependent measure was participants' ratings of their *comfort* during each of the discussions. Similarly, for the analyses of the judges' impressions, dependent measures were judges' perceptions of the participants' honesty, comfort, actual liking for the paintings, and degree of liking that they seemed to be trying to convey. We computed exaggeration scores by subtracting perceptions of actual liking from perceptions of conveyed liking.

[1] Sex of participant also was included as a factor in the design, but the results of that factor are not of central relevance to the theme of the present report and therefore are not included. They are currently available from Bella M. DePaulo and will be reported in a subsequent article that will include several studies in addition to the data from this research (Witt, Bell, & DePaulo, 1996). The significant effects for participant sex that did occur in the present research generally indicated that the overall effects were characteristic of both the men and the women, but they were even more characteristic of the women. For example, the judges believed that both the men and the women were trying to convey more liking for the special paintings than for the not-special ones, but they saw a bigger difference for the women than for the men. The degree-of-investment factor was included to test whether our predictions for investment would be qualified by degree of investment. Although those results will not be presented, significant interactions did occur for participants' self-reports and judges' impressions. In all instances, the interactions indicated that the effects of investment were even stronger when the art students were highly invested in the paintings (the paintings were their own work) than when they were moderately invested in them (the paintings were their favorites). Complete results are available from Bella M. DePaulo.

Participants' Self-Reports and Judges' Impressions

Paintings that were liked and disliked, special and not special. As we predicted, the main effect of investment was significant for all relevant dependent measures (see Table 1 for statistical tests and significance levels). When the paintings were special to the art students (second and fifth columns of Table 1), compared to when they were not (first and fourth columns), the participants reported being more uncomfortable and more dishonest. They also tried to convey more liking, and they exaggerated their liking more. Similarly, all main effects of liking for the painting were significant. Participants said they were less comfortable and less honest when discussing the paintings they disliked than the ones they liked. They tried to convey more liking for the paintings they liked, but they exaggerated their liking more for the paintings they disliked (i.e., they tried to convey more liking than they really did feel). In fact, according to their self-reports, participants actually understated their liking for the liked paintings. Also as predicted, the effects of the artists' investment on participants' honesty and comfort depended significantly on whether the participants liked the paintings. Participants were significantly less honest and less comfortable when discussing the special paintings than the not-special ones only when they disliked the paintings.

The judges also thought that the participants were more uncomfortable and dishonest when discussing the special paintings than the not-special ones and that they tried to convey more liking, and more exaggerated liking, for the special paintings. The differences in honesty and conveyed liking that they noted were even more striking when the participants disliked the paintings than when they liked them.

There was a significant main effect of investment, but no significant interaction with liking for the painting, on judges' impressions of participants' actual liking for the paintings. Participants liked the special paintings almost exactly the same as the not-special ones. The judges did not know this, and from watching the tapes, their impression was that the participants really did like the special paintings even more than the not-special ones.[2]

There was one other way in which the judges' impressions departed from the participants' self-reports. The participants said they exaggerated their liking for the disliked paintings but understated their liking for the liked paintings. The judges thought that the participants were always exaggerating their liking (especially so for the special paintings). Tests of whether the exaggeration (or understatement) scores differed from zero were significant for all four paintings for the participants' self-reports (all $ps \leq .05$ or smaller) and the judges' impressions (all $ps < .001$).

Finally, although the judges were not told whether the participants liked the paintings, their impressions of the discussions of the liked and disliked paintings were accurate. They thought the participants really did like the liked paintings more, and were trying to convey more liking for them, and that they felt less comfortable discussing the disliked paintings.

Honesty and politeness. Did the participants who were instructed to be honest or to be polite behave and feel differently than those who were left to their own devices? Significant main effects of the instructional manipulation for the measures of honesty, $F(2, 82) = 8.42$, $p < .001$, $MSE = 6.92$, and exaggeration, $F(2, 82) = 4.66$, $p = .01$, $MSE = 4.89$, indicated that they did. The means for self-reported honesty in the honest, no-instructions, and polite conditions, were 7.71, 7.85, and 6.95, respectively. The difference between the honest and the no-instructions conditions was not significant. The difference between the no-instructions and the polite conditions was significant, $F(1, 82) = 14.31$, $p < .001$. In their reports of their own honesty, then, participants given no special instructions were more similar to the participants instructed to be honest than to those who were urged to be polite. The judges' impressions of the participants' honesty showed the same thing, $F(2, 82) = 3.03$, $p = .05$, $MSE = 0.85$. The judges thought that the participants were no less honest in the no-instructions condition ($M = 6.49$) than in the honest condition ($M = 6.46$, $F < 1$), but they thought the participants were significantly less honest in the polite condition ($M = 6.30$) than in the no-instructions condition, $F(1, 82) = 5.33$, $p = .02$.

However, in the degree to which they reported exaggerating their liking for the paintings, participants in the no-instructions condition were more similar to participants who were told to be polite. (The uninstructed participants did not differ significantly from the polite participants [$F < 1$], but they did differ significantly from the honest participants, $F[1, 82] = 4.30$, $p = .04$.) In fact, participants in both the no-instructions and the polite conditions said that they tried to convey more liking than they really did feel (Ms = 0.37 and 0.61 for the no-instructions and polite conditions, respectively), but participants in the

[2] We thought that if our judges had instead been completely unaware of the most important constraint in the present research—when the participants were and were not talking to artists who cared—they might have been even more taken by participants' expressions of liking. To test this, we prepared exact typed transcripts of the four conversations of 8 of the participants in the no-instructions and polite conditions who were talking to the artists about paintings that were or were not the artists' own. We recruited 65 raters (32 men and 33 women) to report their impressions of how much the participants really did like the paintings in each conversation, on the same 9-point scale used by our judges. Approximately half of the raters ($n = 33$) rated the conversations with the same information that our judges had—that is, they knew when the artists claimed that the paintings were their own. For the other raters (randomly assigned), that critical information was removed from the transcripts. The key interaction between whether the paintings were or were not special, and whether the judges knew that they were special, was significant, $F(1, 61) = 33.54$, $p < .001$, $MSE = 0.94$. When the paintings were not special, raters perceived almost exactly the same amount of liking when they knew that they were not special ($M = 4.75$) as when they did not know that ($M = 4.79$). However, when the paintings were special and the raters knew that they were, they thought that the participants liked those paintings much *less* ($M = 4.66$) than when they did not have that information ($M = 5.40$). That is, raters discounted some of the liking that participants expressed when they knew that the participants were talking to an artist who cared. The implication for understanding the ratings made by our judges, who *did* know when the paintings were special, is that they may have (inaccurately) perceived even greater differences in liking between the special and not-special paintings if they had *not* had that crucial information.

Table 2
Effects of Instructions and Artists' Investment on Participants' and Judges' Ratings

		Instructions								
		Honest			No instructions			Polite		
Ratings	*Fs*(1, 82)	Not special	Special	Difference	Not special	Special	Difference	Not special	Special	Difference
Participants										
Honesty	7.42****	7.86	7.56	0.30	8.09	7.59	0.50**	7.52	6.39	1.13****
Conveyed liking	6.44***	4.73	4.97	0.24	5.07	5.42	0.35	4.62	5.73	1.11****
Exaggeration	8.58****	−0.25	−0.17	0.08	0.21	0.53	0.32***	−0.02	1.23	1.25****
Judges										
Actual liking	2.96*	4.86	4.84	−0.02	4.77	4.86	0.09	4.73	5.04	0.31***
Conveyed liking	6.30***	5.36	5.50	0.14	5.23	5.64	0.41****	5.24	5.88	0.64****
Exaggeration	3.72**	0.50	0.66	0.16***	0.46	0.78	0.32****	0.51	0.84	0.33****

Note. *Difference* is *not special* minus *special* for honesty, *special* minus *not special* for conveyed liking, exaggeration, and actual liking.
* $p \leq .10$. ** $p \leq .05$. *** $p \leq .01$. **** $p \leq .001$.

honest condition said that they conveyed slightly less liking than they felt ($M = -0.21$).

Did the instructional manipulation influence the way the participants discussed the paintings that were or were not special to the art students? Significant interactions between the instructional manipulation and the investment variable for the measures of honesty, conveyed liking, and exaggeration, indicated that it did. As shown in Table 2, in all three instructional conditions, participants said that they were less honest when the artists were invested in the paintings than when they were not, and they also said that they tried to convey more liking and that they exaggerated their liking more when the artists were invested. The degree to which they showed these effects, however, increased from the honest to the no-instructions to the polite condition. (See the columns in Table 2 labeled *Difference*.) For the exaggeration measure, for example, the degree to which participants exaggerated their liking more for the special than for the not-special paintings was only 0.08 (and not significant) in the honest condition; it increased to 0.32 in the no-instructions condition and to 1.25 in the polite condition. In fact, for all three measures, participants in the no-instructions condition were more similar to the participants in the honest condition than they were to the participants in the polite condition. Contrast analyses showed that the difference between the special paintings and the not-special ones was the same for the honest condition and the no-instructions condition for all three measures ($Fs < 1$); but the special versus not-special difference was significantly greater in the polite condition than in the no-instructions condition for all measures (all $ps = .007$ or smaller).

The judges also thought that the ways that the participants handled the discussions of the special (compared to the not-special) paintings were influenced by their attempts to be honest or polite. The judges thought that the participants tried to convey more liking and more exaggerated liking for the special paintings than for the not-special ones (and they tended to think that the participants really did like the special paintings more, which they did not), and they also noticed that the degree to which participants tried to favor the special paintings increased from the honest to the no-instructions to the polite condition (see Table 2).

The way that the uninstructed participants compared to the others was different for the judges' ratings than for the participants' own reports. In the self-report data, the degree to which the participants favored the special over the not-special paintings was essentially the same for the participants who were told to be honest as for those who were left to their own devices—contrary to our predictions. The judges, in contrast, thought that uninstructed participants were no different from the polite participants in the degree to which they favored the special paintings. (The Fs were < 1 for exaggeration, and 2.52 and 2.51, both $ps = .12$, for actual and conveyed liking, respectively.) The judges also thought that the uninstructed participants were different from the honest participants in the degree to which they favored the special paintings; for conveyed liking, $F(1, 82) = 3.46$, $p = .07$, and for exaggeration, $F(1, 82) = 4.82$, $p = .03$.

Finally, the instructional manipulation was especially important to the way the participants dealt with the artists' investment when the paintings were ones the participants disliked. The three-way interaction of instructions, investment, and liking for the painting was significant for participants' reports of their honesty, $F(2, 82) = 12.23$, $p < .001$, $MSE = 1.59$. As shown in Table 3, when participants liked the paintings (see last three columns of the table), the instructions they received had virtually no effect in any of the conditions on how honest they were about the special compared to the not-special paintings. However, when participants disliked the paintings (first three columns of Table 3), they admitted to being less honest about the ones that were special to the artists compared to the ones that were not. This difference was significant in every instructional condition, but it increased from the honest ($M = 0.50$) to the no-instructions ($M = 0.94$) to the polite condition ($M = 2.38$). Once again, the uninstructed participants, in their self-reports, were more similar to the participants who were told to be honest than to those who were told to be polite. The difference between special and not special was not significantly greater in the uninstructed condition than in the honest condition, $F(1, 82) =$

Table 3
Effects of Instructions, Participants' Liking for the Paintings, and Artists' Investment on Participants' and Judges' Ratings

Ratings and instructions	Type of painting: Disliked: Not special	Disliked: Special	Disliked: Difference	Liked: Not special	Liked: Special	Liked: Difference
Participants: Honesty						
Honest	7.58	7.08	0.50**	8.14	8.04	0.10
No instructions	7.90	6.96	0.94****	8.28	8.22	0.06
Polite	7.48	5.10	2.38****	7.55	7.67	−0.12
Judges						
Honesty						
Honest	6.52	6.39	0.13	6.54	6.37	0.17*
No instructions	6.67	6.31	0.36****	6.58	6.42	0.16*
Polite	6.49	5.97	0.52****	6.42	6.33	0.09
Conveyed liking						
Honest	4.27	4.30	0.03	6.46	6.71	0.25*
No instructions	4.01	4.59	0.58****	6.45	6.69	0.24*
Polite	4.15	5.07	0.92****	6.32	6.70	0.38***

Note. The special paintings were the ones in which the artists were invested. *Difference* is *not special* minus *special* for honesty and *special* minus *not special* for conveyed liking.
* $p \leq .10$. ** $p \leq .05$. *** $p \leq .01$. **** $p \leq .001$.

1.89, $p = .17$, but it was significantly greater in the polite condition than in the uninstructed condition, $F(1, 82) = 20.25$, $p < .001$.

The judges also noticed that it was especially difficult for the participants to discuss the special paintings truthfully when the participants disliked the paintings but were trying to be polite about them, $F(2, 82) = 3.27$, $p = .04$, $MSE = 0.28$. As shown in Table 3, the degree to which the participants seemed to be more dishonest when discussing the disliked paintings that were special to the artists (compared to the disliked paintings that were not special) increased from the honest to the no-instructions to the polite condition. The same pattern occurred for judges' perceptions of the liking that participants seemed to be trying to convey, $F(2, 82) = 4.74$, $p = .01$, $MSE = 0.54$. When the participants disliked the paintings, the judges thought that they seemed to be trying to convey especially more liking for the special than for the not-special paintings and that this tendency increased from the honest to the no-instructions to the polite condition. Again, the judges, in contrast to the participants, thought that the uninstructed participants were more similar to the participants who were told to be polite than to the participants told to be honest. For perceptions of honesty, the special versus not-special difference for the disliked paintings was not significantly smaller in the no-instructions condition than in the polite condition, $F(1, 82) = 1.57$, *ns*, but it was nearly significantly greater in the no-instructions condition than in the honest condition, $F(1, 82) = 3.25$, $p = .07$. For the measure of conveyed liking, the corresponding values were $F(1, 82) = 3.03$, $p = .08$, and $F(1, 82) = 7.93$, $p = .006$.

Verbal Strategies

Design. The design for the analyses of participants' verbal strategies was the same as for the participants' self-reports and the judges' ratings, except that one within-subjects factor of professed affect (liking–disliking) was added. For the measure of explicit evaluation, the levels were (a) whether the participants explicitly said that they liked the painting and (b) whether they explicitly said that they disliked it (see Cochran, 1950, and Rosenthal & Rosnow, 1991, for the use of ANOVA with dichotomous dependent variables). For total aspects mentioned and for new aspects mentioned, the levels were number of liked aspects mentioned and number of disliked aspects mentioned.

Initial likes and dislikes. To be sure that participants did not like more aspects of the special paintings than the not-special ones even before they met the art students, we analyzed the number of liked and disliked aspects of all of the paintings that participants had described in writing. The interaction of investment with number of liked versus disliked aspects was not significant ($F < 1$). Thus, participants began by listing almost exactly the same number of likes and dislikes for the special paintings as for the not-special ones.

Professed affect. The main effect of professed affect was significant for all three measures. Across the discussions of all of the paintings, the participants were almost twice as likely to say that they liked a painting ($M = 0.44$) than to say that they disliked it ($M = 0.24$), even though all participants actually liked the exact same number of paintings that they disliked, $F(1, 82) = 56.09$, $p < .001$, $MSE = 0.13$. They also mentioned many more things that they liked than disliked about the paintings ($Ms = 5.26$ and 3.61), $F(1, 82) = 66.49$, $p < .001$, $MSE = 7.64$, and of the aspects of the paintings that they mentioned but had not originally listed, significantly more of them were aspects that they liked than disliked ($Ms = 3.26$ and 2.04), $F(1, 82) = 43.01$, $p < .001$, $MSE = 6.47$.

Liked and disliked paintings. Participants had more diffi-

culty communicating truthfully about the paintings that they disliked than about the ones that they liked. The interactions between professed affect and liking for the paintings were significant for all three measures. As shown in Table 4, when participants liked a painting, they said so 81% of the time; however, when they disliked a painting, they said so explicitly only 48% of the time, $F(1, 82) = 353.70$, $p < .001$, $MSE = 0.19$. When participants liked a painting, they mentioned many more things about it that they liked than that they disliked, but when they disliked a painting, they mentioned fewer than one more thing about it that they disliked than liked, $F(1, 82) = 217.25$, $p < .001$, $MSE = 5.05$. Similarly, when discussing a painting that they liked, participants mentioned 4.18 additional things about it that they liked that they had not already listed, compared to only 1.33 new things that they disliked; in contrast, the number of new liked and disliked aspects that participants generated when the painting was disliked hardly differed (2.34 and 2.75), $F(1, 82) = 114.39$, $p < .001$, $MSE = 4.33$.

Stonewalling was indicated by the main effect for liking for the painting for the measures of explicit evaluation, $F(1, 82) = 30.37$, $p < .001$, $MSE = 0.09$, and total number of aspects mentioned, $F(1, 82) = 7.87$, $p = .007$, $MSE = 4.09$. These results showed that participants not only had a hard time telling the truth about the disliked paintings, but they also had a hard time saying anything at all. When participants disliked a painting, they were less likely to make any explicit evaluation (whether positive or negative) than when they liked it (Ms = 0.28 and 0.41 for disliked and liked paintings, respectively). Participants also mentioned fewer things that they liked or disliked when they disliked a painting ($M = 4.23$) than when they liked it ($M = 4.64$).

Not-special and special paintings. Professed affect interacted significantly with investment, and in the predicted direction, for all three measures. As shown in Table 5, when the paintings were special to the art students (compared to when they were not), the participants were relatively more likely to say that they liked them and relatively less likely to say that they disliked them, $F(1, 82) = 7.83$, $p = .006$, $MSE = 0.13$. Similarly, when the paintings were special, compared to when they were not, the participants mentioned relatively more things that they liked about them and relatively fewer things that they disliked, $F(1, 82) = 13.14$, $p < .001$, $MSE = 4.22$. Similarly, the participants thought of relatively more new things to like about the special paintings than about the not-special ones, and relatively fewer things to dislike, $F(1, 82) = 5.70$, $p < .05$, $MSE = 3.55$. Thus, the ways in which the participants discussed the special versus the not-special paintings paralleled the ways they discussed the liked versus disliked paintings. It is important to note that the interactions of professed affect with investment were not qualified by participants' liking for the painting (except for one higher order interaction involving instructions, discussed next). That means, for example, that participants used their strategy of mentioning many more liked than disliked aspects (of the special paintings) just as much when discussing the paintings they disliked as the paintings they liked.

Table 4
Participants' Verbal Strategies Used in Discussing the Disliked and Liked Paintings

Verbal strategy	Participants' liking for the paintings: Disliked	Liked	Difference
Explicit evaluation[a]			
Liked	0.08	0.81	0.73****
Disliked	0.48	0.00	−0.48****
Total aspects mentioned			
Liked	3.84	6.67	2.83****
Disliked	4.62	2.61	−2.01****
New aspects mentioned			
Liked	2.34	4.18	1.84****
Disliked	2.75	1.33	−1.42****

[a] Proportion of participants who explicitly said that they liked or disliked the paintings.
**** $p \leq .001$.

Table 5
Participants' Verbal Strategies Used in Discussing the Not-Special and Special Paintings

Verbal strategy	Artists' investment: Not special	Special	Difference
Explicit evaluation[a]			
Liked	0.41	0.48	0.07**
Disliked	0.28	0.21	−0.07**
Total aspects mentioned			
Liked	5.02	5.49	0.47**
Disliked	3.92	3.30	−0.62***
New aspects mentioned			
Liked	3.09	3.42	0.33*
Disliked	2.20	1.88	−0.32

Note. The special paintings were the ones in which the artists were invested.
[a] Proportion of participants who explicitly said that they liked or disliked the paintings.
* $p \leq .10$. ** $p \leq .05$. *** $p \leq .01$.

Honesty and politeness. A main effect of instructions on explicit evaluations indicated that the more polite the participants were instructed to be, the less likely they were to offer any explicit evaluation at all, $F(2, 82) = 3.30$, $p < .05$, $MSE = 0.10$. The means for the honest, no-instructions, and polite conditions were 0.38, 0.35, and 0.30, respectively. The instructional manipulation also moderated the way the participants explicitly evaluated the paintings that were or were not special to the art students. There was a significant interaction among the instructional manipulation, investment, liking for the painting, and professed affect, $F(2, 82) = 4.84$, $p = .01$ (see Table 6). If participants were being completely honest, then they would explicitly say that they disliked the paintings that they actually did dislike and that they liked the paintings that they actually did like. That is, the numbers in the middle two columns of Table 6 would all be exactly 1.00. But none of them were. The numbers were fairly high for the liked paintings; when participants really did like the paintings, between 69% and 91% of them explicitly said that they did. And virtually none of them ever said that

Table 6
Effects of Instructions, Participants' Liking for the Paintings, and Artists' Investment on Participants' Explicit Evaluations

Instructions	Disliked paintings		Liked paintings	
	Professed liking	Professed disliking	Professed liking	Professed disliking
Honest				
Not special	.09	.56	.81	.03
Special	.03	.62	.91	.00
Difference	−.06	.06	.10	−.03
No instructions				
Not special	.00	.64	.83	.00
Special	.16	.40	.79	.00
Difference	.16**	−.24****	−.04	.00
Polite				
Not special	.03	.47	.69	.00
Special	.16	.22	.81	.00
Difference	.13*	−.25****	.12*	.00

Note. The special paintings were the ones in which the artists were invested. Entries are proportions of participants who explicitly said that they liked or disliked the paintings.
* $p \leq .10$. ** $p \leq .05$. **** $p \leq .001$.

they disliked any of those paintings. Neither the artists' investment in the paintings nor the instructions the participants had been given made much of a difference. But when participants disliked the paintings, they often refrained from saying so explicitly, and both the instructions and the artists' investment mattered to them.

As shown in Table 6, participants strayed farthest from the truth when they disliked a painting that was special to the artist and they were trying to be polite about it. When instructed to be honest, 62% of the participants explicitly acknowledged that they disliked the painting that was special to the artist; among the uninstructed participants, only 40% did so, and among those participants urged to be polite, only 22% did so. (All differences among these three numbers were significant [$ps <$ .01 or smaller].) In the no-instructions and polite conditions, 16% of the participants told outright lies: They explicitly said that they liked the painting that they had already indicated in writing that they hated. (In the honest condition, 3% of the participants did this.)

It is also informative to compare the relative percentages of participants who explicitly said that they liked and disliked the detested special paintings in each condition. In the honest condition, 59% more of the participants said that they disliked than liked the painting that they did in fact dislike. In the no-instructions condition the difference was 24%, and in the polite condition it was only 6%.

Participants' explicit evaluations of the disliked paintings that were *not* special to the artists showed that participants used the predicted strategy of evaluation by implication. Participants were somewhat less likely to say explicitly that they disliked the disliked paintings when they were instructed to be polite than when they were instructed to be honest, but this drop from honest to polite was far less precipitous when the paintings were not special to the artist (56 to 47) than when they were special (62 to 22). The converse occurred for explicit statements of liking—the outright lies. Participants in the polite condition (and the no-instructions condition) told polite lies about the paintings they disliked that were special to the art student: 16% of them said that they liked those paintings, compared to 3% in the honest condition. In contrast, when the disliked paintings were *not* special to the artist, 3% of the participants in the polite condition (and none in the no-instructions condition) explicitly said that they liked them, compared to 9% in the honest condition.

In sum, then, when participants were trapped in the challenging situation of trying to be polite about work they disliked that was special to the artists with whom they were interacting, they manipulated both their evaluation of the work in which the artists were invested *and* their evaluation of the other artwork in which the artists had no investment. They refrained from saying explicitly that they disliked the paintings that were special to the artists. At the same time, they were much less restrained when it came to condemning the paintings that were not special.[3]

Discussion

A Looking Glass or a Reversible Figure?

Decades of research relevant to the reflected appraisal process have indicated that our perceptions of others' views of us are not strongly related to their actual views and that our self-perceptions are more highly related to our perceptions of others' appraisals than to their actual appraisals (Felson, 1992). We began with two possible explanations for the poor fit between actual and perceived appraisals. First, people may not be open and honest in communicating their appraisals. Second, we may misperceive those appraisals.

The present study strongly supported the first explanation. The participants refrained from saying how they really did feel about the paintings, especially when they disliked them. By itself, this finding is hardly new. From the literatures on performance appraisals (Fisher, 1979; Larson, 1984, 1986, 1989), the MUM effect (Tesser & Rosen, 1975), and lying in everyday life (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996; see also Folkes, 1982), as well as the literatures that followed more directly from the symbolic interactionist tradition (Blumberg,

[3] An example of a truthful answer to the question "What do you think of it?" was given by a participant discussing a disliked painting that was one of the artist's favorites: "It's ugly. It's just ugly." An example of a truthful message about a liked painting that was one of the artist's favorites was: "I liked it. This was, this was my second favorite of the group. Um, it was the, the detail that was put into, uh, some of the, you know, the, the nuances in color, the way the black is done. And um, and it was, uh, yeah, I really liked it overall." An example of an answer that was coded as a lie (i.e., the participant claimed to like a disliked painting) was given by a participant discussing a painting that was the artist's own work: "I like this one." All participants had more to say about each painting when asked additional questions, but these were their complete answers to the artist's first question ("What do you think of it?"). Over the entire course of the discussion of each painting, participants spoke an average of 217 words.

1972; Felson, 1992; Swann, Stein-Seroussi, & McNulty, 1992), we already knew that there are formidable barriers to the direct communication of appraisals—especially negative ones—to the persons they concern. What our work has shown that *is* new is (a) the powerful impact on appraisals of the target person's personal investment in the object of the appraisals, (b) the difficulty of eliciting totally honest evaluations, and (c) the value of the defensibility postulate in predicting the verbal strategies people use in dodging the truth. Perhaps even more important, our findings suggest that (d) when we look to others for their appraisals, what we see is neither a looking glass nor a hopelessly distorted image, but a reversible figure.

Truth and investment. Among the many motives that have been postulated to account for the reluctance to convey negative evaluations, concern with the target person's feelings is perhaps the one that is most consistently cited and supported. If the target person's feelings are most important, then the more the target person cares about the object of evaluation, the less likely that person should be to hear a truthful appraisal, especially when the truth would hurt. This is so, we predicted, even though the instrumental value of honest appraisals should also increase with the target person's personal investment. Our study was the first to manipulate target persons' personal investment, and our findings were strongly supportive of our predictions. Across virtually every measure, communications were more dishonest when the target persons cared about the objects of the appraisals than when they did not.

Is the truth ever told to those who care? We predicted that, when left to their own devices, people are practitioners of politeness (P. Brown & Levinson, 1987). We are the first to test people's strategic use of polite dissembling by directly instructing some of the participants to behave as politeness theory predicts they would and then comparing their behavior to that of participants given no special instructions. We thought that the uninstructed participants would convey appraisals of the paintings the art students cared about that were just as distorted and dishonest as those conveyed by the participants who were explicitly instructed to be polite and avoid hurting the art student's feelings. According to the judges' perceptions, this is what usually happened. The participants' reports, however, were often at odds with our prediction. The self-reports of the uninstructed participants were usually more similar to those of the participants urged to be honest than to those urged to be polite.

We are inclined to trust the judges' perceptions. The participants in the uninstructed condition may have been motivated to describe themselves as honest. (The polite participants, in contrast, had an excuse for being dishonest—they were following instructions.) The judges had no investment in perceiving the participants as either honest or dishonest, and they made their ratings without any awareness of the participants' instructional conditions. We also trust the judges' perceptions more because they were more in line with the results of our objective measures of what the participants actually said. The results were clearest for our measure of outright lies about the disliked special paintings. The percentage of participants who explicitly said that they liked the paintings that they had already told us that they detested was identical in the no-instructions and the polite conditions (16%); in the honest condition it was lower (3%). The same pattern is evident in the percentage of participants who refrained from saying explicitly that they disliked the detested paintings when they were special compared to when they were not special. In the honest condition, this withholding of an explicit negative evaluation was equally likely for the special as for the not-special paintings, but in the uninstructed and the polite conditions participants were significantly more likely to refrain from saying that they disliked the detested painting when it was special than when it was not. Furthermore, the magnitude of this difference between the special and not-special paintings was virtually identical for the uninstructed and polite participants (Table 6).

In that the participants who were urged to be honest told virtually no lies about the special paintings they disliked, and were no more likely to withhold their explicit negative evaluations of the special paintings than of the not-special ones, were they evenhanded in their discussions of the special and not-special paintings in every other way, too? If so, that would indicate that there is an easy way to elicit totally honest feedback—urge others to tell the truth and give them a compelling reason for doing so (e.g., it is only by hearing totally honest reactions that art students can learn how others really do perceive particular paintings). According to the participants' self-reports, they usually were evenhanded. The one important exception occurred when they were describing the paintings they disliked; in that condition they admitted that they were significantly less honest about the special paintings than the not-special ones (Table 3). The judges, too, thought that the honest participants were usually just as honest when discussing the special paintings as they were when discussing the not-special ones. But again, there was an important exception. The judges thought that the honest participants exaggerated their liking more when they were discussing the special paintings than the not-special ones (see Table 2).

Another condition in which it may have been possible for all participants to be just as honest about the special paintings as the not-special ones was when they liked the paintings. According to their self-reports, participants were in fact evenhanded in their discussions of the special and not-special paintings when they liked the paintings. On no measure did they report significantly less truthfulness. The judges, however, did think there were some differences. For example, they thought that the participants were trying to convey significantly more liking, and that they were exaggerating their liking more, for the paintings they liked that were special to the artists than for the liked paintings that were not special (see Table 1). The objective measures of what the participants really did say lend support to the judges' views. The strategy of amassing positive evidence preferentially for the special paintings (relative to the not-special ones) was just as evident when the participants liked the paintings as when they disliked them. In this study, then, there was essentially no condition under which art students who cared about the paintings heard totally honest feedback about them.

The defensibility postulate. When participants give a painting one of the lowest possible ratings on the liking scale and then tell the art student that they like that painting, it is hard for them to defend that statement as truthful. For that reason, our defensibility postulate predicted that outright lies would occur infrequently, as in fact they did. But they also occurred exactly

when we expected them to—when the participants were discussing paintings they disliked that were special to the art student, and especially when they were given no special instructions or were instructed to be polite.

The strategy of amassing misleadingly positive evidence when one's true opinion is negative is one that was noted in passing nearly four decades ago in a study of the self-restraint of friends (Mayer, 1957). In the present context, participants practiced this strategy by mentioning many aspects of the special paintings that they really did like while mentioning relatively few aspects that they actually disliked. The resulting communications are highly defensible in that the positive qualities that the participants mentioned were ones that they really did like. Although participants were not equally forthcoming about the qualities they disliked, they did not deny disliking those qualities (which would not be defensible); they simply refrained from mentioning them.

The mentioning of new positive qualities that were not initially listed is a riskier strategy, but one that perhaps can work if participants convince themselves that they really do like these newly discovered aspects that they simply had not noticed previously. Credibility is added if the participants also notice some new aspects of the paintings that they dislike, though our prediction is that they will discover fewer of these new disliked aspects than liked aspects when the paintings are special. The results supported that prediction, too.

We believe that future research will show that the strategy of amassing misleading evidence is widely used. Most objects of evaluation—for example, personalities, appearances, job performances, paintings, and journal articles—are complex stimuli that routinely elicit both positive and negative reactions. It is a fairly simple matter, then, when put on the spot to voice one's opinion, to reel off one positive comment after another.

Also as predicted by the defensibility postulate, participants used the very clever strategy of evaluation by implication. By explicitly stating their disliking for the paintings created by other artists, while refraining from stating their disliking for the art student's own work, they implied a favorable social comparison. They appeared to like the art student's own work more than the other artists' work. They never exactly said that, however, so their communications can be defended as truthful.

In contexts in which it is possible for evaluators simply to avoid communicating their appraisals, we think that they will often do just that. In the performance appraisal literature, for example, it has been noted that supervisors sometimes delay giving negative feedback (Larson, 1989). Even when complete avoidance is no longer possible, evaluators still manage to convey less than the whole truth. For example, both supervisors (Larson, 1986) and football coaches (Felson, 1981) hedge by conveying specific appraisals rather than global ones. In the present research, we found that participants stonewalled by offering fewer explicitly evaluative comments and mentioning fewer aspects of the paintings that they liked or disliked when they were discussing paintings they disliked than ones they liked.

The reversible figure. Our results suggest an unanticipated answer to the question of why our perceptions of others' appraisals are not strongly related to others' actual appraisals: Participants described the paintings in ways that allowed the art students a choice as to what to hear and what to believe.

When discussing disliked paintings that the artists cared about, participants exaggerated their liking, withheld explicit expressions of disliking, and even told some outright lies. This gave the art students the opportunity to think that the participants really did like those paintings. But the participants also dropped some blatant hints as to their *relative* degrees of liking for the different paintings that the artists cared about. For example, they did not even *try* to convey as much liking for the paintings they disliked as for the ones they liked. They rated the disliked special paintings a 2 on the 9-point scale at the beginning of the study, and they tried to convey a rating of 4 to the art students. Although this was substantially higher than the liking they really did feel, it was still significantly lower than the degree of liking they tried to convey for the special paintings they really did like—a 7.

By biasing their appraisals of the special paintings in a positive way—by mentioning relatively more things that they liked about them and relatively fewer things that they disliked about them, relative to the not-special paintings—the participants again handed the artists the option of believing that they really did like their paintings. Still, the ratio of liked to disliked aspects that the participants communicated was not as lopsided as when they were describing paintings that they really did like.

Participants' explicit evaluations also allowed for interpretive flexibility. Sixteen percent of the uninstructed and polite participants explicitly said that they liked the special paintings that they actually hated. However, that number was dramatically lower than the 80% of the uninstructed and polite participants who explicitly said that they liked the special paintings that they really did like.

One of the most important qualifications of the symbolic interactionist model that has emerged from research is that the link from metaperception to self-perception is not unidirectional. Although it is true, as the symbolic interactionists have long maintained, that our perceptions of how others view us can influence our self-concept, it is also true that our self-concept can influence our beliefs about how others view us (Felson, 1992; Jussim et al., 1992; McNulty & Swann, 1994). The present research shows how the latter effect might occur. The feedback that people receive, even in the very difficult situations like the ones we created in this research, is unlikely to be totally dishonest and univalent. Instead, it is complex and multifaceted, offering plausible evidence for very different interpretations. This leaves lots of room for self-perceptions to influence the interpretation that is selected.

Qualifications

Our results are qualified in three ways. First, we recruited only women as artists, to keep the size of the study manageable. What was not known then, but is known now, is that people tell more lies to protect the other person's feelings when they are talking to women than to men (DePaulo et al., 1996). If we had included male artists, the overall rate of lying would probably have been lower. But we think that the key relationship between

truth and investment would have remained unaffected. Still, that is a question for future research.

Second, we argued that participants were being less than truthful when they were more effusive about the special paintings than the not-special ones, after indicating in their initial ratings of the paintings that they liked the special and not-special paintings just the same. However, it is possible that over the course of the conversations, participants changed their minds about the special paintings and really did come to like them more. However, even if the process were one of genuine attitude change rather than strategic dissembling, the consequences for the artists remain the same: They heard evaluations of the paintings they cared about that were strikingly more positive than they would have heard if the paintings were not special to them. They could not count on gleaning equally glowing appraisals from evaluators who did not know, or did not care, whether the paintings were special to them.

In fact, although we have argued that lies are told to those who care, we must also acknowledge that sometimes brutal truths are told instead. In our paradigm, this did not occur. Perhaps participants would have told the truth if they cared more about the long-term consequences for the artists of hearing misleading feedback, rather than the short-term consequences of hurting the artists' feelings and feeling badly for having done so. Some evaluators are in roles that demand that they pay attention to the long-term consequences. Pre-med advisors, for example, are duty bound to warn their low-achieving advisees that their career plans may be unrealistic. Still, we suspect they would often find a way to do so politely. Perhaps evaluators will be harshly critical when they are intellectually insecure (Amabile, 1983). But even this effect, we think, is likely to occur only when the evaluators are either protected by anonymity or when they do not need to deliver their feedback to the person they are criticizing in a face-to-face interaction. We think that naked truths will be told when the conventions of polite society have not yet been fully internalized (as in the story of the emperor's new clothes), or when they have been temporarily abandoned, as when adults are caught in the throes of anger or hatred. Other possibilities are better left to research than to speculation.

References

Amabile, T. M. (1983). Brilliant but cruel: Perceptions of negative evaluators. *Journal of Experimental Social Psychology, 19,* 146–156.

Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin, 112,* 461–484.

Bavelas, J. B., Black, A., Chovil, N., & Mullett, J. (1990). *Equivocal communication.* Newbury Park, CA: Sage.

Blumberg, H. H. (1972). Communication of interpersonal evaluations. *Journal of Personality and Social Psychology, 23,* 157–162.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage.* Cambridge, England: Cambridge University Press.

Brown, R., & Gilman, A. (1989). Politeness theory and Shakespeare's four major tragedies. *Language in Society, 18,* 159–212.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika, 37,* 256–266.

Cooley, C. H. (1902). *Human nature and the social order.* New York: Scribner's.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111,* 203–243.

DePaulo, B. M. (1994). Spotting lies: Can humans learn to do better? *Current Directions in Psychological Science, 3,* 83–86.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. E. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70,* 979–995.

DePaulo, B. M., Kenny, D. A., Hoover, C., Webb, W., & Oliver, P. (1987). Accuracy of person perception: Do people know what kinds of impressions they convey? *Journal of Personality and Social Psychology, 52,* 303–315.

DePaulo, B. M., & Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. Yuille (Ed.), *Credibility assessment* (pp. 51–70). Norwell, MA: Kluwer Academic.

DePaulo, B. M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. *Journal of Personality and Social Psychology, 45,* 1096–1103.

DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985a). Deceiving and detecting deceit. In B. R. Schlenker (Ed.), *The self and social life* (pp. 323–370). New York: McGraw-Hill.

DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985b). Telling ingratiating lies: Effects of target sex and target attractiveness on verbal and nonverbal deceptive success. *Journal of Personality and Social Psychology, 48,* 1191–1203.

Felson, R. B. (1981). Self and reflected appraisals among football players: A test of the Meadian hypothesis. *Social Psychology Quarterly, 44,* 116–126.

Felson, R. B. (1992). Coming to see ourselves: Social sources of self-appraisals. In E. J. Lawler, B. Markovsky, C. Ridgeway, & H. A. Walker (Eds.), *Advances in group processes* (Vol. 9, pp. 185–205). Greenwich, CT: JAI Press.

Fisher, C. D. (1979). Transmission of positive and negative feedback to subordinates: A laboratory investigation. *Journal of Applied Psychology, 64,* 533–540.

Folkes, V. S. (1982). Communicating the reasons for social rejection. *Journal of Experimental Social Psychology, 18,* 235–252.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117,* 21–38.

Glaser, B. G., & Strauss, A. L. (1964). Awareness contexts and social interaction. *American Sociological Review, 29,* 669–679.

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior.* Garden City, NJ: Anchor.

Goffman, E. (1971). *Relations in public.* New York: Harper & Row.

Holtgraves, T. (1992). The linguistic realization of face management: Implications for language production and comprehension, person perception, and cross-cultural communication. *Social Psychology Quarterly, 55,* 141–159.

Jones, E. E. (1990). *Interpersonal perception.* New York: Freeman.

Jussim, L., Soffin, S., Brown, R., Ley, J., & Kohlhepp, K. (1992). Understanding reactions to feedback by integrating ideas from symbolic interactionism and cognitive evaluation theory. *Journal of Personality and Social Psychology, 62,* 402–421.

Kenny, D. A., & DePaulo, B. M. (1993). Do we know how others view us? An empirical and theoretical account. *Psychological Bulletin, 114,* 145–161.

Larson, J. R., Jr. (1984). The performance feedback process: A preliminary model. *Organizational Behavior and Human Performance, 33,* 42–76.

Larson, J. R., Jr. (1986). Supervisors' performance feedback to subordinates: The impact of subordinate performance valence and outcome dependence. *Organizational Behavior and Human Decision Processes, 37,* 391–408.

Larson, J. R., Jr. (1989). The dynamic interplay between employees' feedback-seeking strategies and supervisors' delivery of performance feedback. *Academy of Management Review, 14,* 408–422.

Mayer, J. E. (1957). The self-restraint of friends: A mechanism in family transition. *Social Forces, 35,* 230–238.

McNulty, S. E., & Swann, W. B., Jr. (1994). Identity negotiation in roommate relationships: The self as architect and consequence of social reality. *Journal of Personality and Social Psychology, 67,* 1012–1023.

Mead, G. H. (1934). *Mind, self, and society.* Chicago: University of Chicago Press.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research* (2nd ed.). New York: McGraw-Hill.

Schlenker, B. R. (1980). *Impression management.* Monterey, CA: Brooks/Cole.

Snyder, M. L., & Frankel, A. (1976). Observer bias: A stringent test of behavior engulfing the field. *Journal of Personality and Social Psychology, 34,* 857–864.

Swann, W. B., Jr., Stein-Seroussi, A., & McNulty, S. E. (1992). Outcasts in a white lie society: The enigmatic worlds of people with negative self-conceptions. *Journal of Personality and Social Psychology, 62,* 618–624.

Tesser, A., & Rosen, S. (1975). The reluctance to transmit bad news. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8, pp. 193–232). New York: Academic Press.

Witt, C. L., Bell, K. L., & DePaulo, B. M. (1996). *Sex differences in deceptiveness: Real or perceived?* Unpublished manuscript.

Received December 1, 1994
Revision received April 8, 1996
Accepted April 16, 1996 ■

New Editors Appointed, 1998–2003

The Publications and Communications Board of the American Psychological Association announces the appointment of five new editors for 6-year terms beginning in 1998.

As of January 1, 1997, manuscripts should be directed as follows:

- For the *Journal of Experimental Psychology: Animal Behavior Processes,* submit manuscripts to Mark E. Bouton, PhD, Department of Psychology, University of Vermont, Burlington, VT 05405-0134.
- For the *Journal of Family Psychology,* submit manuscripts to Ross D. Parke, PhD, Department of Psychology and Center for Family Studies–075, 1419 Life Sciences, University of California, Riverside, CA 92521-0426.
- For the Personality Processes and Individual Differences section of the *Journal of Personality and Social Psychology,* submit manuscripts to Ed Diener, PhD, Department of Psychology, University of Illinois, 603 East Daniel, Champaign, IL 61820.
- For *Psychological Assessment,* submit manuscripts to Stephen N. Haynes, PhD, Department of Psychology, University of Hawaii, 2430 Campus Road, Honolulu, HI 96822.
- For *Psychology and Aging,* submit manuscripts to Leah L. Light, PhD, Pitzer College, 1050 North Mills Avenue, Claremont, CA 91711-6110.

Manuscript submission patterns make the precise date of completion of the 1997 volumes uncertain. Current editors, Stewart H. Hulse, PhD; Ronald F. Levant, EdD; Russell G. Geen, PhD; James N. Butcher, PhD; and Timothy A. Salthouse, PhD, respectively, will receive and consider manuscripts until December 31, 1996. Should 1997 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 1998 volumes.