Evaluating One Performance Among Others: The Influence of Rank and Degree of Exposure to Comparison Referents

John R. Chambers and Paul D. Windschitl

Pers Soc Psychol Bull 2009; 35; 776 originally published online Mar 23, 2009;
DOI: 10.1177/0146167209333044

Evaluating One Performance Among Others: The Influence of Rank and Degree of Exposure to Comparison Referents

John R. Chambers
University of Florida

Paul D. Windschitl
University of Iowa

Three studies examined the influence of comparison-referent exposure (i.e., the frequency with which one views comparison referents) on evaluations of the ability of a target person (either oneself or another person). In Experiment 1, participants performed a task and then viewed performances of both upward and downward referents. Participants who saw more performances by the upward referents than the downward referents evaluated their own performances less favorably did than participants who saw more performances by the downward referents than the upward referents. Experiment 2 produced similar findings, showing that comparison exposure also influences people's evaluations about someone other than themselves. In Experiment 3, comparison-exposure effects were significantly reduced when participants were instructed to think deliberatively about the comparison information, consistent with the idea that people typically rely on imprecise representations of comparison information even when they are capable of forming more precise representations from memory if motivated to do so.

Keywords: *social comparison; comparative judgments; ability evaluations*

Imagine that you and a friend are approaching a dance studio for your first salsa lesson. Before entering the studio, you gaze through a window and catch glimpses of various dancers as they move in and out of your view. Some of the dancers are clearly quite good, and others are not. Finally, your friend nudges you into the studio, where the two of you are expected to join in and do what you can. How would your observations of dancers through the window influence your confidence and perception of your own salsa skills? And how would they influence your perception of your friend's skills?

Based on a host of previous social comparison research, it seems safe to conclude that watching numerous upward comparison referents (dancers clearly better than you) would result in lower confidence and worse self-perceptions than would watching numerous downward comparison referents (see Festinger, 1954; Kruglanski & Mayseless; 1990; Suls & Wheeler, 2000). In other words, your *rank* among these dancers would matter to you: the better your rank, the more confident and skilled you would feel. Presumably, the same would be true of your perceptions of your friend's skills; the better he or she ranks among the dancers you viewed, the better your impression of his or her skills.

A host of additional research in social cognition provides reasons to expect that aside from rank, the *degree of exposure* to specific social comparison referents could influence your confidence and perceptions (see Higgins, 1996; Wyer & Srull, 1986, 1989). *Degree of exposure* is a general term but could include number of exposures, salience of exposure, and duration of exposure—all factors that have been shown to influence the impact of stimuli, such as a prime or contextual stimulus, on various responses (e.g., Higgins, Bargh, & Lombardi, 1985; Srull & Wyer, 1979; see review by Higgins, 1996). We might reasonably assume, therefore, that the effect of a

Authors' Note: We are grateful to Jerry Suls for comments on an earlier draft of this article. Correspondence concerning this article should be addressed to John R. Chambers, Department of Psychology, University of Florida, PO Box 112250, Gainesville, FL 32608; e-mail: jrchamb@ufl.edu.

PSPB, Vol. 35 No. 6, June 2009 776-792
DOI: 10.1177/0146167209333044

particular dancer on your self-perceptions (or your perceptions of your friend) might be stronger when that dancer passed into your view from the window more times rather than fewer times. In short then, previous research and theory provide good reason to expect that both rank and degree of exposure would play a role in determining how witnessing several other dancers would affect your confidence and self-perceptions (or perceptions of another person).

However, we know of no social comparison or social perception research that has expressly examined the influence of rank and degree of exposure within the same context. Hence, no research has examined the following: Do people typically track rank-order standing, such that the impact of multiple comparison referents (e.g., numerous other dancers) on perceptions is driven primarily by actual rank and not substantially biased by degree of exposure? Alternatively, do people rarely track their specific rank in a given context and instead formulate only general perceptions of how they fall relative to a group of comparison referents?—an approach that would seemingly leave room for degree of exposure to have a notable impact on self-perceptions. The present research addresses such questions.

More specifically, we tested three interrelated predictions in this article. First, we tested whether differential exposure to various comparison referents does indeed have a notable impact on ability or performance evaluations for a target (self or another person). Second, we tested whether degree of exposure can have an influence on evaluations that is independent of perceived differences in rank. Third, we tested whether the relative influence of rank differences and degree of exposure depends on the perceiver's processing goals. Before describing our specific experiments, we first provide additional background regarding our predictions.

ENUMERATED AND NONENUMERATED FREQUENCY ESTIMATION STRATEGIES

We assume that people believe that the numeric rank of the target (self or another person) among comparison referents is an important metric and should be a key basis for ability evaluations. However, we suggest that in many situations, people do not explicitly and precisely generate the value of the target's numeric rank en route to forming an evaluation of the target's ability. In other words, rank may be widely recognized as an important value in judging ability, but it is often not explicitly calculated when judging ability. Furthermore, we suggest that a person will not always generate an accurate rank value for the target even when the person has the requisite episodic knowledge to do so. The person may develop a vague representation of rank, but the vagueness of the representation allows for rank-irrelevant factors—such as degree of exposure—to influence the representation. Returning to the dancing example, if you viewed four dancers, you might be capable of precisely ranking yourself among the four, but we suspect you would not explicitly do so. Instead, you would develop a vague representation of four dancers and your place among them. Because of the vagueness, your self-evaluations would be better if you spent more time watching the two inferior dancers than the two superior dancers, and worse if you spent more time watching the two superior dancers than the two inferior dancers.

Research on frequency estimation strategies provides a backdrop for our arguments (e.g., see Brown, 1995, 1997; Greene, 1986; Hasher & Zacks, 1984; Hintzman, 1988; Manis, Shedler, Jonides, & Nelson, 1993). The basic process of frequency estimation would presumably be a mediating process for making precise numeric rank judgments regarding ability (because you need to determine the number of people who are superior or inferior to the target person). Brown (1995) described several strategies for the basic processes of frequency estimation, and these strategies can be categorized as either enumeration or nonenumeration strategies. In enumeration strategies, people recall instances from memory and count them. For example, to determine the number of faculty members in your area of your department, you might attempt to recall and count the relevant names. Nonenumeration strategies do not involve explicit counting. These strategies produce estimates or qualitative evaluations of event frequency, not precisely accurate frequency counts. One example of a nonenumeration strategy involves probing episodic memory, then *estimating* a frequency on the basis of whether the probe resembles many or few traces (e.g., see Hintzman, 1988). Another example is using ease of recall as a basis for estimating which of two types of events is more frequent (Tversky & Kahneman, 1974).

Enumeration strategies, which can lead to precisely accurate responses under some conditions, generally take longer and require more effort than do nonenumeration strategies, and people may be reluctant to employ enumeration strategies under various conditions (Brown, 1995, 1997). When people rely on nonenumeration strategies such as those mentioned above, their responses are bound to be influenced by frequency-irrelevant factors such as recency effects, event salience, depth of processing, and other biasing factors relevant to episodic memory (see e.g., Hintzman, 1988; Maki & Ostby, 1987; Tversky & Kahneman, 1974).

Now let us connect this work on frequency estimation to our studies, in which people view multiple comparison referents—with the degree of exposure to specific referents being manipulated. Our argument is that people are generally content with formulating only a rough impression of the target person's rank among a group of relevant referents.

Given that people will rely on nonenumerated impressions rather than on enumerated assessments, their impressions of the referents—and therefore the impact that those impressions have on evaluations of a target—will be susceptible to various biasing factors (Hintzman, 1988; Maki & Ostby, 1987; Tversky & Kahneman, 1974; see also Higgins, 1996). The key biasing factor we explore here is degree of exposure; referents seen more often (or for a greater number of repetitions) will have greater impact than referents seen less often (or only once).

Although we are proposing that people's evaluative judgments are typically influenced by the degree of exposure to superior or inferior comparison referents, we are not suggesting that people's enumerated assessments of how many referents the target person outperformed (numeric rank) would be similarly biased, if they were asked to make such assessments. In fact, we assume that the predicted degree-of-exposure effects on evaluative judgments can occur even when any exposure effects on enumerated estimations have been controlled.

EXPERIMENT 1

Experiment 1 focused specifically on a self-judgment context in which the participant made judgments about his or her own ability. The main goal was to test the predictions that degree of exposure to comparison referents can have a notable influence on self-evaluations of ability (at least as much influence as a modest rank manipulation) and can do so independently of any influence exposure has on subjective estimates of rank-order standing.

Participants, who were tested individually, completed a fictitious age-detection test and received bogus performance feedback. They then viewed one of three videotapes of other students (i.e., the comparison referents) doing the same age-detection test before completing the dependent measures questionnaire. By manipulating the content of the tapes, we were able to test for two separate effects. The *comparison-exposure effect* was the extent to which participants' self-evaluations were influenced by their degree of exposure—via the tapes—to superior and inferior referents. The *comparison-rank effect* was the extent to which participants' self-evaluations were influenced by the rank of their own performances (relative to the referents' performances).

Method

Participants. Participants were 138 undergraduate students recruited in partial fulfillment of a research component of an elementary psychology course. The available participant pool contained mostly female students, and because we wanted to examine same-gender comparisons to minimize the potential influence of performance-related gender stereotypes, we tested only women in this experiment.

Procedure. Each participant was told that she would be taking an "age-detection test" that was related to "general competencies at perceiving social cues and navigating a social environment." This test consisted of two sets of 20 cards, each displaying pictures of four adult faces that appeared roughly similar in age. The participant was told that her goal was to select the oldest individual shown on each card and to place the card in one of two piles depending on whether the experimenter indicated her selection was correct or incorrect.

The participant then completed the first of the two card sets and received bogus feedback indicating she had guessed correctly on exactly 10 out of the 20 cards. The experimenter explained that she would next complete a second set of the test, which was "slightly different from the first set, but typically yields scores that are highly related to performance on the first set of the test." The participant then completed the second test set and received bogus feedback indicating she had guessed correctly on 11 out of 20 cards.

The experimenter then told the participant that the next task would be to watch a videotape and then make ratings about unspecified aspects of this videotape. The participant was told that the individuals shown in the videotape were participants in the same experiment from a past semester who had been videotaped completing the same age-detection test and that she would watch various randomly selected segments of these performances. In actuality, all of the so-called participants in the videotape were undergraduate female confederates. The participant watched one of three videotapes (described below) before completing a questionnaire containing all of the dependent measures for this experiment.

Videotapes. In the videotapes, a male experimenter and female confederates were seen engaging in the same age-detection test that the participant had just completed. In all conditions, the videotape showed performance footage from six separate test sets. The footage included only the last six card trials of these sets as well as the conclusions of these sets when the experimenter provided oral feedback about the total number of correct guesses on a given set. The videotapes were shortened in this manner for the purpose of avoiding participant fatigue.

Table 1 provides a detailed representation of the confederates (and their performances) contained on the videotapes used in Conditions A, B, and C. The tape used for Condition A contained performance footage of four

TABLE 1: Performance Scores for Confederates Shown on Videotape in Conditions A, B, and C in Experiments 1 and 2

	Videotape Condition		
	A	*B*	*C*
Confederate 1	15, 15	15	15
Confederate 2	14, 12	13	13
Confederate 3	8	9, 7	9
Confederate 4	6	6, 6	7
Confederate 5			6
Confederate 6			6
Average score across performances	11.67	9.33	9.33
Average score across performers	10.5	10.5	9.33
Rank of participant in relation to confederates shown in videotape	3rd of 5	3rd of 5	3rd of 7

NOTE: Each number in the top six rows reflects the score a confederate received (on videotape) for a given set of the age-detection test. These performances of the confederates were shown in an apparently nonsystematic order, with the two performances of a given confederate (when relevant) always being separated by the performance of another confederate. The bogus scores received by the participant (in Experiment 1) and Sarah (in Experiment 2) were 10 and 11.

separate confederates. Two confederates were better than the participant (scoring above the participant's own scores of 10 and 11), and two were worse. However, the videotape showed two performance sets for each of the upward referents but only one set for each of the downward referents. For example, as indicated in Table 1, the video showed Confederate 2 score 14 on one performance set and 12 on another. The tape used for Condition B also contained footage of four separate confederates. Again, two confederates were better than the participant, and two were worse. However, the videotape showed only one performance set for each of the upward referents but two sets for each of the downward referents. The tape used for Condition C contained footage of six separate confederates. Two confederates were better than the participant, but four confederates were worse. As indicated in Table 1, the videotape for Condition C showed only one set for each confederate.

The comparison between Conditions A and B constitutes a test of the comparison-exposure effect. The videotapes in both of these conditions show two upward and two downward referents. However, for Condition A, the upward referents are seen twice as much as the downward referents, and the reverse is true for Condition B. The comparison between Conditions B and C constitutes a test of the comparison-rank effect. The tape for Condition B shows two upward and two downward referents, whereas the tape for Condition C shows two upward and four downward referents. In other words, whereas the rank-order standing of the participant relative to the confederates is third out of five in Condition B, it is third out of seven in Condition C. Despite this difference, the qualities of the specific performances seen by participants in Conditions B and C were held constant (see Table 1). It is important to emphasize that, given the nature of the study design, the rank of the participants was held constant when testing the comparison-exposure effect, and the degree of exposure to superior and inferior performances was held constant when testing the comparison-rank effect.

Finally, it is important to note that we selected confederates who were not similar to each other in appearance. This was done to ensure that, when participants were viewing a second performance from a given confederate, they would not mistakenly believe it was a new confederate. Also, the specific performances were shown in a fixed but seemingly random manner such that no two performances for a single confederate appeared consecutively within a videotape. We also interspersed the extreme performances on the videotape to avoid undue influence from possible primacy or recency effects.

Dependent-measures questionnaire. A list of the dependent measures is provided in Appendix A. Most of the measures inquired about the participants' evaluations of their own performances or the performances of the other students shown in the tapes. More specifically, the items asked participants how good they were at detecting people's ages, how many students out of 100 they would outperform on the age-detection test, how their age-detection skills compared with those of the students shown in the videotape, how skillful those students were, how satisfied they were with their own performances on the test, and how watching the test performances of the other students made them feel about their own test performances. In addition to these evaluation measures, recall measures asked participants how many unique comparison referents appeared in the videotapes and how many of those referents had outperformed them on the test (i.e., upward referents).[1]

Results

Preliminary analyses. Table 2 displays the means for all the dependent variables. To determine the underlying structure of the responses, we first subjected responses to these seven evaluative measures to a principal axis analysis with varimax rotation. Three interpretable factors emerged (all with eigenvalues exceeding 1), but

TABLE 2: Mean Ratings and Standard Deviations Across Experimental Conditions in Experiment 1

	Videotape Conditions		
Dependent Measures	*A* *M* (*SD*)	*B* *M* (*SD*)	*C* *M* (*SD*)
Evaluation measures			
Skill in relation to comparison referents (index). (Items 4–6)[a]	–0.75 (0.63)	0.24 (0.63)	0.55 (0.66)
How did watching the performances of the other participants affect the way you felt about your performance on the test? (Item 4)[ab]	–0.66 (0.91)	0.10 (0.91)	0.60 (0.76)
Your skill level at detecting age in comparison to the other participants shown in the videotape. (Item 5)[a]	–0.74 (0.95)	0.28 (0.72)	0.52 (0.83)
How good were the people you saw on the videotape at detecting people's ages? (Item 6)[a]	0.81 (0.73)	–0.34 (0.76)	–0.53 (0.93)
How good are you at detecting people's ages? (Item 2)	0.13 (0.85)	0.10 (1.06)	–0.25 (1.06)
How satisfied are you with your performance on the age-detection test? (Item 1))	–0.13 (1.02)	0.01 (0.94)	0.13 (1.04)
If 100 other college students took the same age-detection test, how many would score lower than you did on the test? (Item 3)	–0.21 (0.85)	–0.07 (0.99)	0.31 (1.10)
Your general social skills and competencies compared to the social skills and competencies of other students. (Item 7)	0.20 (1.13)	–0.02 (0.98)	–0.19 (0.83)
Recall items			
How many people shown in the videotape performed better than you did on the age-detection task? (Item 8)	2.21 (0.46)	2.04 (0.55)	2.19 (0.72)
How many different people were shown in the videotape (not including the experimenter)? (Item 9)[ab]	3.88 (0.53)	4.17 (0.48)	5.98 (1.14)

NOTE: Values reported for the evaluation measures are standardized scores. Item numbers correspond to the items in the dependent measures questionnaire (see Appendix A). Dependent measures superscripted with the letter *a* indicate differences between the Videotape A and B conditions significant at the .05 level. Dependent measures superscripted with the letter *b* indicated differences between the Videotape B and C conditions significant at the .05 level.

only the first factor contained more than one item. This first factor, which may roughly be called "skill in relation to comparison referents," included the three items asking participants how their age-detection skills compared with the those of the referents shown in the video tape, how watching the performances of the referents affected their feelings about their own performances, and how good the referents shown in the videotape were at the test (reverse keyed). Responses to these three items were centered before averaging (coefficient $\alpha = .79$), with a higher score on this index essentially suggesting that a participant evaluated the self favorably and/or the referents unfavorably after viewing the tape. Each of the remaining evaluative items (four of the seven) either loaded on a single-item factor or did not load on any factor (most important, their cross-loadings with the "skill" index were weak, all < .20).

Evaluation measures. When scores on the "skill in relation to comparison referents" index were submitted to an ANOVA, there was a significant effect of videotape condition, $F(2, 134) = 51.58$, $p < .001$, partial $\eta^2 = .44$. Comparisons using a Bonferroni correction were conducted to examine differences among the three conditions. First, we examined the comparison-exposure effect by comparing the responses of participants in Condition A, who watched multiple performances by upward referents, to the responses of Participants in Condition B, who watched multiple performances by downward referents. As expected, the participants who watched multiple performances by downward referents—relative to participants who watched multiple performances by upward referents—believed that the referents were less skillful at age detection, that watching the performances of those referents made them feel better about their own test performances, and that their own performances compared more favorably against those of the referents, $t(92) = 7.65$, $p < .001$. This was the case even though participants' actual rank standings among the referents in the videotapes was the same in both conditions (it was third of five).

Second, we examined the comparison-rank effect by comparing the responses of participants in Condition B, whose scores on the age-detection test ranked third of five in relation to the referents, with the responses of participants in Condition C, whose scores ranked third of seven. Even though participants in these two conditions differed in their precise rank standing among the referents, they did not differ in terms of how they perceived the abilities of those referents or how they thought their own and the referents' performances compared, $t(88) = 2.27$, $p > .05$ (although participants in Condition C said that watching those referents made them feel better than did participants in Condition B, $t(88) = 2.84$, $p < .05$). Thus, unlike the comparison-exposure effect, the comparison-rank effect received very little support.

For the remaining four items not included in the index (i.e., satisfaction with test performance, perceived general ability to detect ages, general age-detection ability compared to the average student's ability, and the number of fellow students participants estimated they would outperform on the test), no significant differences were found between either the A and B conditions or the B and C conditions, all *p*s > .10.

Recall items. A critical question is whether our comparison-exposure manipulation influenced participants' evaluative judgments because they were not attentive to the number of unique upward or downward referents shown in the videotapes and/or they misperceived their actual rank-order standings in relation to those of the referents. Although no differences were detected between Conditions A and B in participants' recall of the number of referents shown with superior scores on the age-detection test, $t(93) = 1.60$, $p > .05$, participants in Condition A did in fact believe that in the videotape there were fewer unique referents than was reported by participants in Condition B, $t(93) = 2.84$, $p < .01$. This difference suggests that participants in the two conditions may have differed in their beliefs about their rank order in relation to the referents, with participants in Condition A assuming that they held less favorable rank-order standings than what participants in Condition B assumed.

Although we thought it was possible that the comparison-exposure manipulation could lead to differences in participants' beliefs about their rank-order standings, we also predicted that comparison exposure could influence people's self-evaluations of ability apart from any influence it has on their perceived rank order. To assess this possibility, we computed a ratio index for each participant by dividing her estimate of the total number of referents receiving better scores than her own score by her estimate of the total number of unique referents on the videotape. Higher scores on this index indicate the participant believed she had a less favorable rank-order standing. We then conducted multiple regression analyses in which ratings for the "skill" index was the dependent variable, and the ratio index scores and videotape condition (A or B) were predictors.

The results of these analyses revealed that, after we controlled for their beliefs about their rank-order standings (i.e., the ratio index), participants who observed multiple performances by upward referents (Condition A) still held less favorable beliefs about their skill in relation to the referents than those beliefs held by participants in Condition B, $\beta = .58$, $t(91) = 6.92$, $p < .001$. Thus, the comparison-exposure effect was at least partially independent of participants' beliefs about their rank-order standings. Interestingly, the ratio index was not a significant independent predictor, $\beta = -.14$, $t(91) = -1.64$, $p > .10$, indicating that perceptions of their exact rank among the referents had less influence on participants' evaluative judgments than did their amount of exposure to specific upward versus downward referents.

That comparison exposure had an independent effect on evaluative judgments is important because it indicates that even when one's perceived rank standing among a group of comparison referents is taken into consideration, differential exposure to upward and downward comparison targets continues to have notable effects on one's self-evaluations. In other words, even among participants who believed their rank among the comparison referents was the same (e.g., third out of five), those who were exposed to multiple performances by downward referents still evaluated their own performances and abilities more favorably than did those who were exposed to multiple performances by upward referents.

As an even more conservative test, we performed a restricted analysis on the subset of participants ($n = 65$, or 70%) in Conditions A and B who correctly recalled *both* the number of upward referents and the total number of unique referents in the videotape. Even among these participants, whose recall of the referents was perfectly accurate, there was still a comparison exposure effect, $t(63) = -6.46$, $p < .001$, with participants reporting that their performances and abilities compared less favorably with the referents' when they were exposed to a greater number of performances by upward referents (Condition A: $M = -0.65$) than by downward referents (Condition B: $M = 0.25$). This finding further bolsters our claim that comparison exposure can have a robust influence on self-evaluative performance judgments, sometimes even overriding knowledge about one's exact rank standing.

Discussion

There were four key findings from Experiment 1. First, participants' self-evaluations were affected by their degree of exposure to upward versus downward referents, such that those who witnessed multiple performances by upward (rather than downward) referents felt their skills were inferior to those of the referents. Second, even when participants' beliefs about their actual rank standings among the referents were statistically controlled, comparison exposure still affected participants' self-evaluations and their perceptions of the referents' abilities. Third, participants' thoughts about how their performances and abilities compared with the referents' were affected by comparison exposure even among participants who were able to recall the number of upward and downward referents perfectly. Fourth, the comparison-exposure manipulation produced a robust effect on the "skill in relation to comparison referents" index, but the comparison-rank manipulation did not.

EXPERIMENT 2

In the introduction, we suggested that comparison exposure may influence self-evaluations because, in most everyday situations, people form rather crude, imprecise representations of comparison information, which would allow the most mentally accessible referents (presumably those seen more often) to have the greatest influence on subsequent ability evaluations. The results of Experiment 1 are fully consistent with this generic cognitive account. However, an alternative explanation for Experiment 1 is that upward and downward social comparisons might have direct affective consequences. Specifically, watching numerous performances by an upward referent (a performer who is obviously better than the observer) may give feelings of frustration and anxiety greater time to build, and watching numerous performances by a downward referent may give one extra time to feel content and perhaps revel in being superior (Brickman & Bulman, 1977; Salovey & Rodin, 1984; Tesser, 1991). The buildup of either positive or negative affect, which would presumably differ in degree across Conditions A and B in Experiment 1, could influence self-evaluations.

Although we find this affective account to be conceivable, and although it may contribute to exposure effects in situations like those faced by participants in Experiment 1, we do not think that affect is a crucial component of exposure effects. In other words, we believe that "cold" cognitive mechanisms (such as enhanced accessibility for referents that have been more frequently observed) can underlie exposure effects even when affective considerations are negligible. This view would be supported if an exposure effect was observed in a paradigm in which people evaluated another person's abilities (rather than their own), where, presumably, minimal affect would be aroused. Therefore, Experiment 2 examined the influence of degree of exposure to referents on participants' evaluations of another person's ability. Doing so can also tell us whether degree of exposure effects generalize beyond self-perception situations to include social evaluations of others.

The design and procedures were very similar to those of Experiment 1. However, instead of completing the age-detection test themselves and evaluating their own performances, participants in this experiment watched the performances of another "participant" in the experiment (actually a confederate) and made judgments about that person's abilities after watching one of the three videotapes that were used in Experiment 1.

Method

Participants. Participants were 129 male and female undergraduate students who participated in partial fulfillment of a research component of an elementary psychology course.

Procedure. The cover story, instructions, and description of the age-detection test were similar to those used in Experiment 1. Participants were told that they would first watch videotaped performances of a "previous participant in this experiment named Sarah" doing an age-detection test and that they should gain a general impression of her performance and ability on the test.

Participants then observed the videotape of Sarah (an undergraduate confederate) completing the last several trials within two sets of the age-detection test and the experimenter providing Sarah with performance feedback indicating that she had scored 10 and then 11 out of 20 correct. The experimenter told participants that they would next be watching a videotape of other students participating in the same experiment completing the same age-detection test and that they should form a general impression of the abilities of those other students. Participants were then randomly assigned to watch one of the three videotapes, which were identical to those shown in Experiment 1 (see Table 1). Afterward, they answered a questionnaire similar to that used in Experiment 1, except the questions referred to Sarah's skills and abilities rather than the participants' own (see Appendix A).

TABLE 3: Mean Ratings and Standard Deviations Across Experimental Conditions in Experiment 2

	Videotape Conditions		
Dependent Measures	*A* *M (SD)*	*B* *M (SD)*	*C* *M (SD)*
Evaluation measures			
Sarah's skill in relation to comparison referents (index). (Items 4–7)[a]	–0.48 (0.70)	0.18 (0.56)	0.37 (0.71)
How did watching the performances of other participants affect the way you felt about Sarah's performance on the test? (Item 4)[a]	–0.49 (0.94)	0.22 (0.94)	0.38 (0.93)
Sarah's skill level at detecting age in comparison to the other participants shown in the videotape. (Item 5)[a]	–0.58 (1.01)	0.44 (0.69)	0.23 (0.95)
How good were the people you saw on the videotape at detecting people's ages? (Item 6)[a]	0.58 (0.90)	–0.10 (0.71)	–0.56 (1.01)
Sarah's general social skills and competencies compared to the social skills and competencies of other students. (Item 7)	–0.28 (0.89)	–0.03 (0.92)	0.34 (1.11)
How good is Sarah at detecting people's ages? (Item 2)	–0.09 (0.89)	0.07 (1.17)	0.04 (0.96)
If 100 other college students took the same age-detection test, how many would score lower than Sarah did on the test? (Item 3)	–0.19 (0.86)	0.07 (1.09)	0.15 (1.04)
Recall items			
How many people shown in the videotape performed better than Sarah did on the age-detection task? (Item 8)[ab]	**2.47 (0.83)**	**1.85 (0.62)**	**2.36 (0.82)**
How many different people were shown in the videotape (not including the experimenter)? (Item 9)[b]	3.96 (0.98)	4.05 (0.60)	5.86 (0.78)

NOTE: Values reported for evaluation measures are standardized scores. Item numbers correspond to the items in the dependent measures questionnaire (see Appendix A). Dependent measures superscripted with the letter *a* indicate differences between the Videotape A and B conditions significant at the .05 level; those with the letter *b* indicate differences between the Videotape B and C conditions significant at the .05 level.

Results

Preliminary analyses. Table 3 displays the means for all the dependent variables. As in Experiment 1, we first submitted responses to the six evaluative measures to a principal axis analysis with varimax rotation to determine their underlying structure. A single factor emerged (with an eigenvalue exceeding 1) that closely resembled that in Experiment 1. This factor may be called "skill in relation to comparison referents," and the four items that loaded highly on it asked participants how Sarah's age-detection skills compared with those of the students shown in the videotape, how watching the performances of those students affected how they felt about Sarah's performance on the test, how Sarah's general abilities compared with other students' general abilities, and how good the other students shown in the videotape were at the test (see Appendix A), with the last item reverse-keyed. These four items were centered before averaging (coefficient $\alpha = .75$), with higher numbers on this index indicating that participants judged Sarah's skills and performance more favorably than those of the other students (referents) shown in the videotape. The factor loading for these four items all exceeded .45, and their cross-loadings with the remaining two items (perceptions of Sarah's general age-detection ability and estimates of the number of students she would outperform on the test) were all lower than .20.

Evaluation measures. When ratings for the "skill in relation to comparison referents" index were submitted to an ANOVA, there was a significant effect of videotape condition, $F(2, 126) = 20.43$, $p < .001$, partial $\eta^2 = .24$. Follow-up comparisons using a Bonferroni correction were next performed to test the comparison-exposure and comparison-rank effects. The comparison-exposure effect was significant, $t(85) = 4.85$, $p < .001$: Participants exposed to multiple performances by upward referents (Condition A) evaluated Sarah's skills and abilities worse in relation to the referents shown in the videotape than did participants exposed to multiple performances by downward referents (Condition B). However, the comparison-rank effect (Condition B vs. C) was not significant, $t(80) = 1.25$, $p > .10$. This overall pattern—a significant exposure effect and nonsignificant rank effect—is the same as was found for Experiment 1.

For the remaining two items that did not load on the primary factor—judgments of Sarah's general age-detection ability and estimates of the number of students she would outperform on the test—there was no effect of videotape condition, F's < 1.36, ps > .10, and follow-up comparisons found no differences between Conditions A and B or between Conditions B and C. So, analogous to the pattern for Experiment 1, differential exposure to upward and downward referents affected thoughts about Sarah's ability and performance in relation to those referents, but not thoughts about her general ability level. We will return to this dissociation between context-specific and global evaluations in the General Discussion.

Recall items. Analyses of our recall items revealed that our comparison-exposure manipulation did not influence participants' beliefs about the total number of unique referents who were shown in the videotape, $t < 1$, although it did affect their estimates of the number of referents who performed better than Sarah did on the age-detection tests (participants in Condition A recalled seeing more upward referents than those in Condition B), $t(85) = 3.87$, $p < .001$. These recall results essentially suggest that participants may have assumed that Sarah held a more favorable rank in Condition B than in Condition A.

Therefore, as was done in Experiment 1, we computed a ratio index in which a higher score indicated that the participant believed that Sarah had a less favorable rank-order standing in relation to the referents. We then conducted regression analyses in which ratings for the "skill in relation to comparison referents" index was the dependent variable and the ratio index scores and condition (A or B) were the predictors. The results of this analysis reveal that, after controlling for their beliefs about Sarah's rank-order standing (i.e., the ratio index), participants who observed multiple performances by upward referents (Condition A) were still more inclined than participants in Condition B to believe that Sarah's skill and ability were inferior to the those of the referents shown in the videotape, condition: $\beta = .23$, $t(84) = 2.33$, $p < .05$. Whereas perceived rank standing was found to have no independent influence on self-evaluations in Experiment 1, Sarah's perceived rank did influence participants' appraisals of her ability in this experiment independent of the comparison-exposure effect, rank index: $\beta = -.49$, $t(84) = -5.10$, $p < .01$. The more crucial finding from Experiment 2, however, is that the comparison-exposure effect was at least partially independent of participants' beliefs about Sarah's actual rank standing among the other students.

To address this issue even further, we identified the subset of participants ($n = 40$, or 31%) who were able to correctly recall the precise number of upward and downward referents shown in the videotapes. Even among these participants, there was still a directional (but nonsignificant) comparison-exposure effect, with those who observed multiple performances by upward referents judging Sarah's relative performance and abilities slightly less favorably than those who observed multiple performances by downward referents (Condition A: $M = 0.87$, $SD = 0.53$; Condition B: $M = 1.10$, $SD = 0.59$), $t(38) = -1.30$, $p < .20$. Obviously, the small sample size for this comparison limits the power to detect a difference, but it is instructive that the effect size was moderately large ($d = .40$).

Discussion

The results from Experiment 2 closely paralleled the main findings from Experiment 1. In particular, comparison exposure had a robust influence on how participants evaluated Sarah's abilities relative to those of the other performers they observed, and this effect was again at least partially independent of changes in perceived rank standing. The comparison-rank manipulation, which did not influence how participants evaluated their own performances and abilities in Experiment 1, also had no influence on how participants in this experiment evaluated Sarah's performance and ability. More generally, these findings show that comparison exposure matters not only in how people judge themselves and their own performances but also in how they perceive the performance and ability of another person, thus broadening the range of comparison situations to which this effect applies.

The results of Experiment 2 also help rule out affect as a valid account of the critical processes behind the exposure effects in Experiments 1 and 2. That is, the fact that participants' evaluations of *another person's* ability shifted as a function of comparison exposure suggests that these exposure effects (of the types detected in Experiments 1 and 2) are unlikely to be critically dependent on strong affective reactions.

EXPERIMENT 3

For Experiment 3, we sought to more specifically investigate the idea that people—when thinking about comparison information—tend to rely on imprecise representations even when they may be capable of generating more precise representations from information available in memory. Recall that we suggested a key reason why comparison exposure matters is because, in most situations, people are not motivated to enumerate referents and to generate precise representations of the numbers of referents with abilities inferior or superior to their own. Rather, we propose that people ordinarily rely on relatively vague representations of comparison information generated with limited effort through—at best—nonenumerated estimation strategies (if any strategy is applied at all). Under such low-effort and nonenumerated processing, information-accessibility differences—such as those caused by differences in degree of exposure to referents—would be especially influential in ability or performance evaluations. Yet, our explanation also assumes that people's episodic memories might harbor enough information for them to generate an assessment of their actual rank standings if required to. This numeric assessment might be biased by the comparison-exposure manipulation, but we suggest that it would be less biased than would evaluative ability judgments and that the bias on evaluative judgments would persist even when the bias in explicit numerical assessments of rank is controlled or absent. The analyses we conducted involving recall measures in Experiments 1 and 2 provide some evidence in support of this view.

However, a better approach to testing our account would involve directly manipulating people's processing goals, to see if people who are encouraged to think numerically and deliberatively about comparison information respond differently to exposure manipulations than do people who are not encouraged in that way. This is precisely what we did in Experiment 3. As in Experiment 2, participants judged the abilities of an unfamiliar target person after first learning about the performances of various upward and downward referents (with the exposure of those performances manipulated). However, immediately before making their ability judgments (but after seeing all the performances), some participants were given very strong instructions to think in careful detail about the performances of the referents, including the numbers of people who did better and worse than the target person. Other participants were not given these instructions. We expected that participants who were not given special instructions would (by default) show robust exposure effects like those in Experiments 1 and 2. In contrast, those given instructions should feel motivated to think more effortfully and precisely about the relative abilities of the referents and to ignore differences in degree of exposure to the referents. Thus, there should be less impact of the exposure manipulation on participants' ability judgments in the latter condition, and if anything, their judgments should become more reflective of the target person's actual rank standing.

We altered the procedures of this experiment in three other important ways. First, whereas we manipulated comparison exposure in Experiments 1 and 2 by sometimes showing *different* performances by the same referent, we manipulated exposure in this experiment by sometimes showing the *same* performance by a referent multiple times. This new manipulation allows us to address a potential alternative explanation. Because participants in Experiments 1 and 2 viewed two separate performances by some referents, it could be argued that they had a more reliable base of knowledge about the true abilities of those referents. If so, then perhaps participants should give more weight to those referents, and this differential weighting would have produced the same pattern of results we observed in those experiments. We do not believe this differential-reliability account provides a viable explanation for Experiments 1 and 2.[2] However, this account needs a direct challenge. If comparison-exposure effects are observed even when the number of unique performances by a referent is held constant, this differential-reliability account would be rendered a less plausible explanation for our findings.

Second, we pitted the comparison-rank and comparison-exposure effects against each other more directly in this experiment. We did this by creating two conditions: in the first condition, fewer upward referents appeared than downward referents, but the performance of each upward referent was repeated twice and the performance of each downward referent was seen only once. In the second condition, this was reversed: More upward referents appeared than downward referents, but the performances of downward referents were repeated twice. In essence, the target person's rank standing was superior in the first condition, but the relative amount of time upward referents appeared was also greater in this condition.[3] Notice that pitting the two effects against each other directly represents a very direct test of whether exposure has an effect on evaluations that can be independent of rank. We expected that—within the specific parameters of the study—the influence of exposure would overwhelm the countervailing influence of rank standing. That is, we predicted a comparison-exposure main effect. We also predicted that participants' judgments would reveal stronger comparison exposure effects in the no-instruction condition than in the careful-instruction condition.

Third, we used a new performance task to strengthen the generalizability of our findings. The target person and referents in this experiment were heard engaging in a memory-recall task, instead of the age-detection test used in the previous experiments.

Method

Participants. Participants were male and female undergraduate elementary psychology students ($N = 341$). They received credit toward a research exposure requirement.

Procedure. Participants were seated at individual computers on which all instructions, manipulations, and dependent measures were presented. Information given to the participants suggested that several students in a previous semester had completed a memory-recall task as part of a class demonstration. It was explained that these students had each been given a short period of time to rehearse and memorize a list of common words (e.g., "moon," "farmer," "tree"), after which the lists were taken away and the students had been given approximately 25 seconds to verbally recall as many of the words from the list as possible. Students' performances on these memory trials had ostensibly been audio recorded (in fact, the recordings were the voices of female undergraduate confederates), and participants were informed that they would be listening and evaluating a small sample of those performances. They were told that each student's word responses in the audio clips would be immediately followed by one of two distinct auditory tones indicating whether the response was correct (a "hit," or a word that appeared on the original list) or incorrect (a "miss," or a word that was not on the original list). Participants learned that because of the random nature in which the computer selected the sample performances from a database, they might hear the performances of some students multiple times.

Next, the performances were played. In this experiment—unlike Experiments 1 and 2—the performance of the target person, called Tracy, was heard *after* the performances of the referents were heard. Participants listened to brief audio clips of seven different referents. As each clip started, the supposed picture, name, and identification number of the referent appeared on the computer screen. We did this to help participants distinguish between the various referents and their performances. As indicated in Table 4, there were three upward and four downward referents in Condition A, and four upward and three downward referents in Condition B. Although varying in exact quality, the upward referent performances in both conditions were relatively superior (i.e., confederate was heard giving many hit and few miss responses), whereas the downward referent performances were relatively inferior (i.e., many miss and few hit responses). In Condition A, upward referent performances were repeated twice (but not consecutively), and downward referent performances were heard only a single time; this was reversed in Condition B. The order of the various upward and downward referents was completely randomized in both conditions. The last person to appear in both conditions was always the target person, Tracy (whom participants were told to pay close attention to); Tracy was heard giving a relatively mediocre test performance (a moderate number of both hits and misses).

After participants heard the audio clips and before they completed the dependent measures, we manipulated the instructional set. By random assignment, some participants were given instructions to think carefully

TABLE 4: Performance Qualities for Confederates Heard in Conditions A and B of Experiment 3

	Condition	
	A	*B*
Confederate 1	Upward (twice)	Upward (once)
Confederate 2	Upward (twice)	Upward (once)
Confederate 3	Upward (twice)	Upward (once)
Confederate 4	Downward (once)	Upward (once)
Confederate 5	Downward (once)	Downward (twice)
Confederate 6	Downward (once)	Downward (twice)
Confederate 7	Downward (once)	Downward (twice)
Proportion of time spent watching upward rather than downward comparison performances	6 to 4	4 to 6
Rank of Tracy in relation to comparison referents	4th of 8	5th of 8

NOTE: *Upward* indicates that the confederate's performance was better than Tracy's; *downward* indicates that it was worse than Tracy's. *Once* and *twice* refer to the number of times the performance was heard by the participant.

and rationally about the referents' performances on the word recall test (careful-instruction condition). These instructions read as follows:

> Please think very carefully about the information you were presented as you are making these ratings. Think logically and rationally about how Tracy's performance on the memory test compared with the other students' performances, much like a scientist would think. That is, try to think about the precise number of students who did better than Tracy on the test versus the number who did worse. Your answers to these questions will be checked for accuracy, so please try to give the most accurate response.

Other participants (no-instruction condition) were not given these instructions and proceeded immediately to the dependent measures after listening to the performance trials. In summary, the experiment was a 2 (audiotape Condition A or B) × 2 (no instruction or careful instruction) between-subjects factorial.

Dependent measures. The primary dependent measures, shown in Appendix B, asked participants how good Tracy's performance was, how many students (out of 100) Tracy would outperform on the test, how Tracy's memory abilities compared with those of the other students on the audiotape, how good the other students were at the memory test, how listening to the

TABLE 5: Mean Ratings and Standard Deviations Across Instruction and Comparison Conditions in Experiment 3

	Instruction Condition			
	No Instruction		*Careful Instruction*	
Dependent Measure	*A* *M* (*SD*)	*B* *M* (*SD*)	*A* *M* (*SD*)	*B* *M* (*SD*)
Evaluation measures				
General performance evaluation (index). (Items 1-6)[ab]	–0.26 (0.72)	0.33 (0.74)	–0.17 (0.77)	0.10 (0.84)
How good was Tracy's performance on the memory-recall test? (Item 1)[a]	–0.27 (0.95)	0.34 (0.87)	–0.10 (0.94)	0.02 (1.14)
If 100 other students completed the memory-recall test, how many would Tracy do better than? (Item 2)[a]	–0.20 (0.84)	0.29 (0.96)	–0.08 (0.97)	–0.01 (0.97)
How do Tracy's general memory abilities compare with those of the other students you heard? (Item 3)[ab]	–0.25 (0.90)	0.31 (0.96)	–0.23 (1.01)	0.15 (1.02)
How good are the other students you heard at the memory-recall test? (Item 4)[ab]	0.21 (0.96)	–0.31 (0.90)	0.31 (1.03)	–0.20 (0.93)
How did listening to the other students' performances affect your thoughts about Tracy's performance? (Item 5)[a]	–0.29 (0.92)	0.36 (0.97)	–0.16 (1.02)	0.09 (0.98)
How did Tracy's performance on the test compare with the performances of the other students you heard? (Item 6)[ab]	–0.32 (0.97)	0.37 (0.96)	–0.19 (0.96)	0.14 (0.98)
Recall items				
Not including Tracy, how many different people did you listen to? (Item 7)	7.48 (5.32)	7.15 (2.31)	6.85 (1.96	7.16 (2.43)
How many of the students you heard performed better than Tracy on the memory-recall test? (Item 8)[a]	3.76 (1.73)	3.06 (1.36)	3.54 (1.62	3.33 (1.39)
Estimate the number of words Tracy correctly recalled from the test list. (Item 9)	5.90 (1.45)	6.32 (1.39)	5.84 (1.50	5.92 (1.84)
Estimate the number of words Tracy incorrectly recalled from the test list. (Item 10)	3.31 (1.00)	3.29 (1.16)	3.35 (1.18	3.61 (1.24)

NOTE: Values reported for evaluation measures are standardized scores. Item numbers correspond to the items in the dependent measures questionnaire (see Appendix B). Dependent measures superscripted with the letter *a* indicate differences between the A and B conditions in the no instruction condition significant at the .05 level; those with the letter *b* indicate differences between the A and B conditions in the careful instruction condition significant at the .05 level.

other students' performances affected their impressions of Tracy's performance, and how Tracy's performance on the test compared with those of the other students. Several recall items were included to check the accuracy of participants' memory for Tracy's and the other students' performances (see Appendix B).

Results

Table 5 displays mean ratings across conditions. As with the first two studies, we subjected responses to the six evaluative measures to a principal axis analysis with varimax rotation to discover the underlying factor structure. Only a single, unitary factor emerged (eigenvalue = 3.92, all others < 1), and all items loaded highly on this factor (all > .43). Thus, after reverse coding the item that asked how good the referents were at the memory-recall test, ratings for the six measures were centered and then averaged to create a "general evaluations" index (α = .89), with higher ratings indicating more favorable evaluations of Tracy's performance and ability.

To test our prediction that comparison exposure would exert less influence on ability judgments when participants were induced to carefully consider the referents and their specific performances, we submitted ratings for the general evaluations index to a 2 (audiotape) × 2 (instruction) between-subjects ANOVA. A significant main effect of audiotape condition, $F(1, 334) = 26.72$, $p < .001$, partial $\eta^2 = .07$, showed that, as expected, participants who observed multiple performances by upward referents (Condition A) judged Tracy's performance and general ability more negatively than did participants who observed multiple performances by downward referents (Condition B). This was true even though Tracy's rank standing among the referents in the audiotape was actually superior in Condition A than in Condition B.

More important, this effect of the exposure manipulation was qualified by the predicted Audiotape × Instruction interaction, $F(1, 334) = 3.64$, $p = .05$, partial $\eta^2 = .01$. As expected, the comparison-exposure effect on participants' judgments was stronger when they were not instructed to think carefully about the referents or their performances and weaker when they were given such instructions: no-instruction condition, $F(1, 164) = 27.20$, $p < .001$, partial $\eta^2 = .14$; and careful-instruction condition, $F(1, 170) = 4.96$, $p = .03$, partial $\eta^2 = .03$—although the effect was still present even in the latter condition.

Recall items. The comparison exposure manipulation affected the number of upward referents participants remembered hearing (*M*s of Condition A vs. B = 3.65 vs. 3.20), $F(1, 335) = 7.43$, $p < .01$, partial $\eta^2 = .02$, but not the number of unique referents they remembered (7.16 vs. 7.16), $F < 1$ (instruction condition had no effect on either item, *F*s < 1). Hence, Tracy's perceived rank (calculated from these recall items) was better in Condition B than in Condition A. This finding is particularly striking because Tracy's true rank among all performers was actually better in Condition A than in Condition B.

Although we expected that comparison exposure could influence perceptions of Tracy's rank (and hence could be a partial mediator of the influence of the exposure manipulation on judgments about Tracy), we also expected it would have an influence on judgments about Tracy independent of any bias in her perceived rank standing.[4] Therefore, we conducted a series of regression analyses like those done in Experiments 1 and 2.

As expected, audiotape condition predicted evaluative judgments *independent* of Tracy's perceived rank. In line with our predictions, this was true of participants in the no-instruction condition but not those in the careful-instruction condition. In the no-instruction condition, for example, participants who heard the downward referents twice (relative to those who heard the upward referents twice) evaluated Tracy's more favorably, regardless of how they perceived her rank on the test: audiotape condition, $\beta = .18$, $t(162) = 3.08$, $p < .01$, and perceived rank, $\beta = -.63$, $t(162) = -10.62$, $p < .001$. In contrast, when given instructions to think carefully about the referents, participants' evaluations of Tracy strictly depended on how they perceived her rank, and not on the degree of exposure they had to upward versus downward referents: audiotape condition, $\beta = .09$, $t(169) = 1.51$, $p > .10$, and perceived rank, $\beta = -.65$, $t(169) = -11.31$, $p < .001$.

Discussion

Experiment 3 builds on our prior experiments in several important respects. First, we showed a comparison exposure effect in a completely different task involving perceptions of memory ability. Second, by repeating the same performances twice rather than showing two different performances by a given referent, we now rule out the differential-reliability account of the exposure effects. Third, we found evidence that comparison-exposure effects can be moderated by the evaluator's processing goals. Namely, when not given any special prompts to think carefully about the comparison information, participants' judgments about the target person's performance were affected by the degree of exposure to upward versus downward referents, but when they were explicitly told to think carefully, their judgments showed the exposure effects to a lesser extent.

Finally, it should be noted that in Experiment 3, the significant exposure effects imply not only that exposure mattered but that it mattered enough to outweigh the influence of rank-order standing. Recall that in terms of rank-order standing, Tracy was slightly better in Condition A than in Condition B, yet people in Condition B evaluated Tracy more favorably because of the exposure differences that were part of the audiotape manipulation.

GENERAL DISCUSSION

There are many everyday circumstances in which people evaluate a target person (either oneself or another person) within the context of several comparison referents. Although previous social-cognitive research (e.g., showing that prime frequency affects judgment; Higgins et al., 1985; Srull & Wyer, 1979) provides reasons to suspect that degree of exposure to specific referents could influence target evaluations, no previous research has specifically tested this possible exposure effect and compared it to rank effects. Our experiments tested both exposure and rank effects within the same experimental context. Consistent with our predictions, we found that differential exposure to upward and downward referents—even when objective rank is controlled—does indeed have a significant impact on performance evaluations.

Readers with a social-cognitive background might be far from surprised that differential exposure—when manipulated separately from rank—matters for self and social perceptions. However, our experiments reveal more than this basic point. First, our experiments demonstrate that exposure effects are not dwarfed in size by rank effects. In Experiments 1 and 2, the effect of the exposure manipulation was stronger than that observed for a modest rank manipulation. In Experiment 3, when exposure and rank effects were pitted against each other in a combined manipulation, participants' evaluations were more strongly influenced by the exposure difference than by the rank difference. We do not wish to claim that exposure effects are typically larger than rank effects in most environments—a conclusion that would seem to be impossible to effectively test—but we do think this work suggests that exposure effects might be just as important to consider as rank effects when trying to understand how comparison referents influence self and social evaluations.

Second, our experiments demonstrate that the influence of exposure (manipulated separately from rank) on evaluations was not fully mediated by subjective rank. Prior to these experiments, one might have expected that exposure manipulations would influence perceptions of rank, which then influence evaluations of the target. Yet we expected that exposure manipulations would influence evaluations even when perceived rank was statistically controlled. Our results are consistent with our argument that people do not base evaluations on perceived rank status per se but instead are typically content to rely on imprecise representations of available comparison information. Given that people rely on nonenumerated impressions rather than on enumerated assessments, their impressions of the referents—and therefore the impact that those impressions have on evaluations of a target—are susceptible to exposure effects.

Third, Experiment 3 demonstrated that the relative influence of degree of exposure and rank differences on performance evaluations depends on the evaluator's processing goals. The degree of exposure had a stronger influence when participants were not instructed to think carefully about the comparison referents. Hence, it appears that the participants' default approach (when given no special instructions) was to rely on a nonenumerated strategy for assessing how the target's performance compared to others'. Yet, participants had the capability of forming more precise (enumerated) representations from memory because when they were urged to think carefully about the comparison referents (after having seen those referents), their evaluations were less influenced by the exposure manipulation.

Finally, these experiments also rule out two alternative accounts of the exposure effects. Experiments 2 and 3 seem to preclude an affective account—which suggests that the buildup of affect upon encountering an upward or downward referent might explain exposure effects. Although this account might be relevant to some instances in which one is making a self-evaluation about an important skill or attribute (see, e.g., Tesser & Campbell, 1982), the account seems implausible for explaining exposure effects when people are evaluating a stranger, such as in Experiment 2 and 3. Experiment 3 also casts doubt on the differential-reliability account, given that participants in Experiment 3 saw only one performance from a given referent (either once or twice). Thus, data from these experiments are clearly consistent with our explanation for exposure effects and inconsistent with at least two other explanations.

Why Did We Only Observe Contrast Effects?

In our studies, participants' evaluations of the self or another target person always showed *contrast effects* after learning the comparison information. That is, they evaluated the self (or Sarah or Tracy) more negatively after extended viewing of upward referents and more positively after extended viewing of downward referents. This probably owes to the particular features of our design, performance task, and comparison referents. In designing our studies, we made efforts to ensure that the referents would be perceived by our participants as similar (e.g., by being college students of the same sex as target person), relevant (i.e., performing the same task), distinctive, and relatively extreme on the evaluative dimension (i.e., performances much better or worse than the target person's). These are some of the factors known to produce reliable contrast effects in judgments (e.g., Mussweiler, Ruter, & Epstude, 2004; Stapel & Koomen, 2000; Stapel & Winkielman, 1998).

Context-Specific Versus Global Evaluations of Ability

One issue that deserves special mention concerns the fact that comparison exposure had stronger effects on some types of evaluation items than on others in Experiments 1 and 2. Namely, the exposure manipulation significantly affected responses to evaluation questions that were either about the referents or that asked about the target's (self or Sarah) skill in relation to the referents' skills. We will call these questions the *context-specific* items. The exposure manipulation (and the comparison-rank manipulation) had no effect on responses to questions that concerned the target's global ability level and did not mention the referents (the *global* items). Also, factor analyses conducted in both studies revealed that the items concerning global ability levels tended to not load on the same factor as the context-specific items.

Interestingly, this dissociation between context-specific and global evaluations has been observed before by other social comparison researchers. Using a self-evaluation paradigm somewhat related to our own, Buckingham and Alicke (2002) had participants take a fictitious lie detection test before receiving bogus performance feedback indicating that they had done better or worse than various referents. After receiving this comparative performance feedback, participants evaluated themselves on both a global ability dimension (e.g., "How would you rate your lie detection ability?") and a context-specific one (e.g., "How well do you think you performed on the lie detection test?"). Most important for our findings, Buckingham and Alicke observed consistent effects of the comparison feedback on participants' context-specific evaluations of ability, but null or inconsistent effects on their global ability evaluations (see also Sanbonmatsu, Harpster, & Akimoto, 1994). They discussed various reasons for this dissociation, including the idea that global ability evaluations are inherently more ambiguous—and hence, more susceptible to biased interpretation by the evaluator (e.g., Dunning, Meyerowitz, & Holzberg, 1989)—than are context-specific evaluations, an idea we believe has some intuitive appeal.

Conclusion

Our research shows that the degree of comparison-referent exposure can influence both self-evaluations and evaluations of another individual. The comparison-exposure effects in our experiments were not trivial in magnitude. These findings should not be taken to suggest that a person's comparison rank is unimportant in his or her self-evaluations (or social evaluations). However, knowing that the comparison exposure effects in our experiments were as strong or stronger than the comparison rank effects is instructive for intuitively gauging the importance of exposure in other social-comparison effects. Returning to the dancing scenario that started this article, although your view from the studio window might tell you that you and your friend are better than half and worse than half of those you watched, your impression of just how skilled the two of you are may depend on whether the window afforded you a more frequent view of the better dancers or the worse ones.

APPENDIX A
THE DEPENDENT MEASURES IN EXPERIMENTS 1 AND 2

Evaluation measures	
1. How satisfied are you with your performance on the age-detection test?	1 = *Very unsatisfied*, 7 = *Very satisfied**
2. How good [are you/is Sarah] at detecting people's ages?	1 = *Not good*, 7 = *Very good*
3. If 100 other college students took the same age-detection test, how many do you think would score lower than [you/Sarah] did on the test?	_____ (numerical estimate)
4. How did watching the performances of other participants affect the way you felt about [your own/Sarah's] performance on the age-detection test?	–3 = *It made me feel much worse*, +3 = *It made me feel much better*
5. Rate [your/Sarah's] skill level at detecting age in comparison to the other participants shown in the videotape. [You were/Sarah was] . . .	–3 = *Much less skillful*, +3 = *Much more skillful*
6. How good were the people you saw on the videotape at detecting people's ages?	1 = *Not good*, 7 = *Very good*
7. Please rate [your/Sarah's] general social skills and competencies compared to the social skills and competencies of other students.	1 = *Much worse*, 7 = *Much better*
Recall items	
8. How many people shown in the videotape performed better than [you/Sarah] did on the age-detection task? _____ (numerical estimate)	
9. How many different people were shown in the videotape (not including the experimenter)? _____ (numerical estimate)	

NOTE: The specific phrasing of the dependent measures in Experiment 1 versus Experiment 2 appears in brackets. The dependent measure with an asterisk did not appear in the questionnaire in Experiment 2.

APPENDIX B
THE DEPENDENT MEASURES IN EXPERIMENT 3

Evaluation measures	
1. How good was Tracy's performance on the memory-recall test?	1 = *Not at all good*, 7 = *Very good*
2. If 100 other students completed the memory-recall test, how many would Tracy do better than?	___ out of 100 total students
3. How do Tracy's general memory abilities compare with those of the other students you heard?	–3 = *Tracy's are much worse*, 0 = *Same*, +3 = *Tracy's are much better*
4. How good are the other students you heard at the memory-recall test?	1 = *Not at all good*, 7 = *Very good*
5. How did listening to the other students' performances affect your thoughts about Tracy's performance?	–3 = *It made me think her performance was much worse*, 0 = *It had no effect*, +3 = *It made me think her performance was much better*
6. How did Tracy's performance on the memory-recall test compare with the performances of the other students you heard?	–3 = *Tracy's was much worse*, 0 = *Same*, +3 = *Tracy's was much better*
Recall Items	
7. Not including Tracy, how many different people did you listen to? Please give your best estimate.	___ different persons
8. How many of the students you heard performed better than Tracy on the memory-recall test? Please give your best estimate.	___ students
9. Please estimate the number of words Tracy correctly recalled from the test list. That is, how many "hits" did she have? Please give your best estimate, even if it is only a guess.	___ words
10. Please estimate the number of words Tracy incorrectly recalled from the test list. That is, how many "errors" did she have? Please give your best estimate, even if it is only a guess.	___ words

NOTES

1. Five additional measures were included but were less germane to the hypotheses being tested (e.g., the perceived validity of the test). All of these measures yielded null results and are not discussed further.

2. The differential-reliability account is implausible for two reasons. First, we told participants in Experiments 1 and 2 that performers tend to receive the same scores on multiple performances of the age-detection test (and in fact, participants in Experiment 1 were provided with nearly identical scores on their successive rounds of the test). Therefore, participants would have little reason to doubt that the performances of the comparison referents shown only a single time in the videotapes were any less reliable than the performances of the comparison referents shown twice in the videotapes. Second, if participants' thoughts about score reliabilities were sophisticated enough to cause them to differentially weight the comparison referents (leading to differences between Conditions A and B), then participants would surely also have been sensitive the comparison-rank manipulation between Conditions B and C. Given that the comparison-rank manipulation had little effect on participants' ability evaluations in either Experiment 1 or 2, the differential-reliability account seems to be an implausible explanation for our findings.

3. Our manipulation of the target person's rank—either fourth or fifth out of eight—in this study addresses a potential limitation in Experiments 1 and 2. In those experiments, rank was manipulated by systematically varying the number of people that the target outperformed (third out of five vs. third out of seven). However, in terms of *nominal* rank, the target was always third. A critic might argue that this is not truly a rank manipulation, especially because there are real-world contexts for which the number of people who rank below a target is not seen as consequential (e.g., a third-place Olympic finisher always gets the bronze medal regardless of the number of competitors). In Experiment 3, even nominal rank is manipulated (fourth or fifth).

4. We were unable to perform a restricted analysis with the subset of participants who correctly recalled the upward and downward because the number of these participants was too small (< 20) to permit any meaningful conclusions.

REFERENCES

Brickman, P., & Bulman, R. J. (1977). Pleasure and pain in social comparison. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 149-186). Washington, DC: Hemisphere.

Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1539-1553.

Brown, N. R. (1997). Context memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 898-914.

Buckingham, J. T., & Alicke, M. D. (2002). The influence of individual versus aggregate social comparison and the presence of others on self-evaluations. *Journal of Personality and Social Psychology, 83,* 1117-1130.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082-1090.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117-140.

Greene, R. L. (1986). Effects of intentionality and strategy on memory for frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 489-495.

Hasher, L. & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39,* 1372-1388.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology handbook of basic principles* (pp. 133-168). New York: Guilford.

Higgins, E. T., Bargh, J. A., & Lombardi, W. (1985). The nature of priming effects on categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 59-69.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528-551.

Kruglanski, A., & Mayseless, O. (1990). Classic and current social comparison research: Expanding the perspective. *Psychological Bulletin, 2,* 195-208.

Maki, R. H., & Ostby, R. S. (1987). Effects of level of processing and rehearsal on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 151-163.

Manis, M., Shedler, J., Jonides, J., & Nelson, T. E. (1993). Availability heuristic in judgments of set size and frequency of occurrence. *Journal of Personality and Social Psychology, 65,* 448-457.

Mussweiler, T., Ruter, K., & Epstude, K. (2004). The man who wasn't there: Subliminal social comparison standards influence self-evaluation. *Journal of Experimental Social Psychology, 40,* 689-696.

Salovey, P., & Rodin, J. (1984). Some antecedents and consequences of social comparison jealousy. *Journal of Personality and Social Psychology, 47,* 780-792.

Sanbonmatsu, D. M., Harpster, L. L., & Akimoto, S. A. (1994). Selectivity in generalizations about self and others from performance. *Personality and Social Psychology Bulletin, 20,* 358-366.

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, 37,* 1660–1672.

Stapel, D. A., & Koomen, W. (2000). Distinctiveness of others, mutability of selves: Their impact on self-evaluations. *Journal of Personality and Social Psychology, 79,* 1068-1087.

Stapel, D. A., & Winkielman, P. (1998). Assimilation and contrast as a function of context-target similarity, distinctness, and dimensional relevance. *Personality and Social Psychology Bulletin, 24,* 634-646.

Suls, J., & Wheeler, L. (2000). *Handbook of social comparison: Theory and research*. New York: Klewer Academic.

Tesser, A. (1991). Emotion in social comparison and reflection processes. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 117-148). Hillsdale, NJ: Lawrence Erlbaum.

Tesser, A., & Campbell, J. (1982). A self-evaluation maintenance approach to school behavior. *Educational Psychologist, 17,* 1-12.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Wyer, R. S., & Srull, T. K. (1986). Human cognition in its social context. *Psychological Review, 93,* 322–359.

Wyer, R. S., & Srull, T. K. (1989). *Memory and cognition in its social context*. Hillsdale, NJ: Lawrence Erlbaum.

Received February 9, 2008
Revision accepted January 17, 2009