BASIC AND APPLIED SOCIAL PSYCHOLOGY, 22(2), 85–100
Copyright © 2000, Lawrence Erlbaum Associates, Inc.

Effects of Category-Based Expectancy Violation on Affect-Related Evaluations: Toward a Comprehensive Model

Cyndi Kernahan
Department of Psychology
University of Wisconsin, River Falls

Bruce D. Bartholow and B. Ann Bettencourt
Department of Psychology
University of Missouri, Columbia

We examined the processes that unfold from category-based expectancy violations. In Experiment 1, which included both Black and White participants, descriptions of in-group and out-group targets' academic credentials either violated or were consistent with relevant category-based expectancies for their respective group category. The results showed that expectancy violations were associated with more extreme affect-related evaluations, regardless of whether the participants were in-group or out-group members. Moreover, these affect-related evaluations were made more quickly than attributions. In Experiment 2, a judgment facilitation paradigm was used in which participants' response times to make affect-related evaluations and attributions regarding expectancy violating targets were compared. Consistent with Experiment 1, the results showed that affect-related evaluations were made more quickly than attributions, and facilitation scores indicated that affect-related evaluations are likely mediators of attributions. Results were consistent with the expectancy-violation theory, and supported the idea that affect-related evaluations mediate the relation between target characteristics and attributional processes.

Expectancy-violation theory (Jackson, Sullivan, & Hodge, 1993; Jussim, Coleman, & Lerch, 1987; also see Burgoon, 1993; Jones, 1990) predicts that persons who violate expectations for their salient group memberships (i.e., violate stereotypes) will be evaluated more extremely than persons who do not. Supporting this premise, studies show that, as compared to targets depicted in ways that are consistent with category-based expectancies, those depicted in ways that violate such expectancies are evaluated more extremely, in the direction of the valence of the target information (e.g., Bettencourt, Dill, Greathouse, Charlton, & Mulholland, 1997; Coleman, Jussim, & Kelly, 1995; Jackson et al., 1993; Jussim et al., 1987; Jussim, Fleming, Coleman, & Kohberger, 1996). For example, in studies examining these effects, a Black person with excellent college or job credentials is often presented as a target that violates category-based expectancies and, in line with expectancy-violation theory, this target is often evaluated extremely positively.

Whereas this initial work revealed reliable effects of category-based expectancy violations on extremity in evaluations, the sequence of the processes that unfold from such violations has yet to be fully investigated. Therefore, the primary purpose of this work was to conduct studies that would provide information about the sequence of processes that follow from category-based expectancy violation. Moreover, although a couple of studies (Bettencourt et al., 1997; Coleman et al., 1995) seem to suggest that the group membership of the evaluator does not moderate effects of category-based expectancy violation, more work is needed to substantiate this finding. Thus, a secondary purpose of this work was to further examine extremity in evaluations by including perceivers from more than one in-group membership. In what follows, we review the relevant literature as well as a process model of category-based expectancy violation proposed by Jackson et al. (1993). In doing so, we suggest an addition to Jackson et al.'s model and also attempt to conceptually integrate a number of theoretical models that

Requests for reprints should be sent to B. Ann Bettencourt, Department of Psychology, 104A McAlester Hall, University of Missouri, Columbia, MO 65211. E-mail: psyannb@showme.missouri.edu

have been used to understand extreme evaluations. Finally, we report the findings of two studies and discuss the implications of their results.

CATEGORY-BASED EXPECTANCY

As we have noted, research provides support for the prediction that category-based expectancy violations induce evaluative extremity (e.g., Bettencourt et al., 1997; Biernat, Vescio, & Billings, 1996; Branscombe, Wann, Noel, & Coleman, 1993; Coleman et al., 1995; Jackson & Cash, 1985; Jackson et al., 1993; Jussim et al., 1987; Jussim et al., 1996). Among these, Jackson et al.'s (1993) pair of studies are the most pertinent for the purposes of this work. In these studies, Jackson et al. provided participants with target information via an undergraduate college application, experimentally manipulated to create conditions that either violated or did not violate respective stereotypes about Black and White students. Jackson et al.'s pretest confirmed that the highly qualified Black college applicant and the poorly qualified White college applicant violated participants' stereotyped expectations. The participants provided "overall evaluation ratings of applicant's overall qualifications and recommended acceptance to the 4-year university" (p. 72). Consistent with expectancy-violation theory, Jackson et al. showed that the well-qualified Black applicant was evaluated more favorably than the similarly qualified White applicant, and that the poorly qualified White applicant was evaluated more negatively than the similarly qualified Black applicant.

Jackson et al.'s (1993) results also showed that participants' attributions about the target applicant were affected by the applicant's group membership. Specifically, category-consistent information tended to be attributed to more stable causes whereas category violations were explained with unstable and external causes. (These effects were found regardless of whether attributions were measured with a rating scale or by participants' self-generated attributions.) Regression and correlational analyses suggested that attributions mediated the relation between the target's racial group and overall evaluations. Based on these findings, Jackson et al. discussed the sequence of processes that unfold from category-based expectancy violations. In their preliminary process model they proposed:

> In this model, a target's behavior is first judged against standards provided by the group stereotype. One of two outcomes is possible from this comparison; either the target's behavior is evaluatively consistent with the group stereotype or it is not. If the target's behavior is consistent with the group stereotype, then attributions to internal causes, namely, ability and effort, will follow. If the target's behavior is inconsistent with the group stereotype, then attribution to external causes or to internal, unstable causes will follow. ... Evaluations, on the other hand, follow in part from attributions. ... Thus, group stereotypes influence target evaluations only indirectly, by way of their influence on attributions. (p. 82)

Other related research (Bettencourt et al., 1997; Biernat et al., 1996) has shown that evaluators' perceptions of the degree of expectancy violation mediate evaluations of targets who vary by ethnic group membership. Therefore, that a target's behavior is first compared to a stereotype associated with a group category has received empirical support. These findings, together with Jackson et al.'s (1993), provide initial support for Jackson et al.'s preliminary model.

However, we propose a key addition to Jackson et al.'s (1993) preliminary process model. In our prior work (Bettencourt et al., 1997), we argued that perceivers are likely to make initial evaluations that ensue directly from the experience of surprise associated with category-based expectancy violations. That evaluations are made relatively effortlessly and spontaneously has been proposed and empirically supported by Bargh and his colleagues (Bargh, Chaiken, Govender, & Pratto, 1992; Chaiken & Bargh, 1993). In addition, expectancy violations, in general, have been shown to be associated with affective responses (e.g., Clary & Tesser, 1983; House & Perney, 1974; MacDowell & Mandler, 1989), and category-based expectancy violations, in particular, have been shown to be associated with mood change (Biernat et al., 1996). In this work, we propose that after the comparison of the target information to the group stereotype (i.e., perception of category-based expectancy violation), but prior to the initiation of attributional processing, perceivers readily make affect-related evaluations; that is, we predict that the processes that unfold on exposure to category-based expectancy violation involve: (a) a comparison to the group stereotype, (b) an affective evaluation, and (c) attributions about the target.

Clearly this prediction is not entirely in line with Jackson et al.'s (1993) proposal that attributions mediate overall evaluations. However, from our perspective, it is likely that perceivers make additional judgments or readjust their initial evaluations of the violating target after they have engaged in the causal attribution process. Indeed, we (Bettencourt, 1997; Kernahan & Bettencourt, 1998) have argued elsewhere that perceivers' motives may induce them to modify their initial evaluations of a violating target. In this work, we concur with Jackson et al. (1993) that attributional processes will affect later judgments about persons who violate category-based expectancies. For the purposes of this article, however, we are only interested in evaluations that are associated with perceivers' initial affective reactions to category-based expectancy violations instead of judgments about the applicants' qualifications or recommendations for college acceptance. We conceptualize these latter judgments measured by Jackson et al. as much less affective in nature (i.e., more cognitive) than the type of evaluations used in this study (e.g., good vs. bad).

Supporting our proposed modification to Jackson et al.'s (1993) model, several lines of research suggest that

attributional explanations may be more effortful than affective evaluations. Bargh et al. (1992) showed that evaluations of good versus bad occur almost instantaneously, and Verplanken, Hofstee, and Janssen (1998) further showed that the affective component of attitudes is often more quickly accessed than the cognitive component. Other research has shown that causal attributions may require effortful processing (Macrae & Shepherd, 1991; Smith & Miller, 1979). For example, in a study using a reaction time paradigm, Smith and Miller (1979) revealed that responses to attribution questions were reliably slower than those to trait inferences and likability ratings. Theoretically, cognitive processes associated with longer reaction times should be those that require more processing steps than those associated with shorter reaction times (Bassili & Racine, 1990; Fiske & Taylor, 1991; Graham & Brown, 1988). Thus, if affect-related evaluations of category-based expectancy violations are more formative in the process than are causal attributions, response times required to make affect-related evaluations should be faster than those to make causal attributions.

It should also be noted that, for those targets whose information does not violate category-based expectancies, attributions may be made relatively quickly (Greenberg & Pyszczynski, 1982; Trope, 1986). When information is consistent with category-based expectancies, "ready-made" attributional explanations associated with stereotypes may be readily accessible. In contrast, when behavior disconfirms a categorical expectancy, effortful causal explanations are likely to be generated in an attempt to account for the disconfirming behavior (Greenberg & Pyszczynski, 1982).

EVALUATORS' IN-GROUP MEMBERSHIP AND OTHER THEORETICAL PERSPECTIVES

Most tests of expectancy-violation theory (e.g., Bettencourt et al., 1997; Branscombe et al., 1993; Jackson et al., 1993; Jussim et al., 1987) have included participants who represent only one group membership. For example, White participants are often asked to evaluate either a Black target or a White target who either does or does not violate category-based expectancies (e.g., Bettencourt et al., 1997; Biernat et al., 1996; Jackson et al., 1993; Jussim et al., 1987). In one study, however, Coleman et al. (1995) included only Black participants who evaluated either a Black or a White job applicant. Interestingly, the results of Coleman et al.'s study replicated the results of a prior study by Jussim et al. (1987) that used identical manipulations and only White participants. The similarity among the Black and the White participants' ratings in these two studies implies that the effects of category-based expectancy violations may induce more extreme evaluations regardless of the group memberships of perceivers. Supporting this implication, the results of another study (Bettencourt et al., 1997) showed that female and male perceivers similarly evaluated female and male job applicants with favorable qualifications more positively when the application violated sex-based stereotypes than when the application did not. Taken together, these findings provide initial evidence that evaluative extremity is associated with category-based expectancy violation, regardless of the perceiver's group membership. In addition, other evidence from our laboratory (Kernahan, Greathouse, & Bettencourt, 1996) has demonstrated that Asian and White participants respond similarly to category-based expectancy violations.

This continuity of evaluative patterns among perceivers from varying group memberships apparently contradicts the predictions made by several other theoretical perspectives (Linville & Jones, 1980; Marques & Yzerbyt, 1988). Among these, both the complexity-extremity hypothesis and the black-sheep hypothesis predict that the in-group membership of the perceiver should play a role in predicting extreme evaluations of individual targets. Important, however, these theoretical perspectives make opposing predictions about which target group member—the in-group or the out-group member—will be evaluated more extremely. Whereas the complexity-extremity hypothesis predicts that out-group members will be evaluated more extremely (i.e., out-group polarization), the black-sheep hypothesis predicts that in-group members will be evaluated more extremely (i.e., in-group polarization). Not surprisingly, the explanatory mechanisms for these two predicted effects are quite different. For example, the complexity-extremity explanation focuses on differences in the complexity of evaluators' cognitive schemas for their in-groups (more complex cognitive schema, less polarization) and out-groups (less complex cognitive schema, more polarization) as responsible for out-group polarization in evaluations. In contrast, the black-sheep hypothesis (Marques, Robalo, & Rocha, 1992; Marques & Yzerbyt, 1988; Marques, Yzerbyt, & Lyens, 1988) focuses on social identity motives as responsible for evaluative extremity in in-group evaluations.

We (Bettencourt et al., 1997) have argued that the apparent contradictions between Linville and Jones's (1980) out-group polarization findings and Marques and colleagues' (Marques et al., 1992; Marques & Yzerbyt, 1988; Marques et al., 1988) in-group polarization findings may be explained, in part, by expectancy-violation theory (also see Branscombe et al., 1993, for a somewhat different explanation). We suggest that, for example, the excellent Black law school applicant in Linville and Jones's (1980) study may have violated category-based expectations for Blacks, and thus this target was evaluated more favorably than the similarly qualified White target. By comparison, we suggested that the unskillful speaker in Marques and Yzerbyt's (1988) study (a study that showed support for the black-sheep hypothesis) may have violated expectations for the in-group (i.e., philosophy major) and thus this in-group target engendered particularly harsh evaluations.

THIS ARTICLE

This set of studies was designed for two main purposes. The primary purpose was to examine our addition to Jackson et al's (1993) sequence model of the processes that unfold from category-based expectancy violation; that is, whether in response to category-based expectancy violations, affect-related evaluations are likely to precede the initiation of attributional processes. Toward this end we measured the time it took participants to key in a number that corresponded to their responses (evaluations, attributions) on numbered rating scales. Although we realized that using a numbered rating scale format would be less sensitive for the purposes of detecting response-time differences than would a yes–no response format (e.g., Smith & Miller, 1979), using numbered rating scales was necessary in Experiment 1 for examining extremity in evaluations. Due to the fact that other research has shown that attention latencies are more reliable with a yes–no response format (e.g., Forgas, 1995), we used this method in Experiment 2.

Our second purpose was to examine whether the effect of expectancy violation on affect-related evaluations is similar across in-group and out-group participants. To our knowledge, this study is the first to explicitly examine whether ethnicity of the participant moderates the effect of ethnic-group expectancy violations. With this in mind, in Experiment 1 we included Black and White participants because these categories are among those most often included as target groups in expectancy violation studies. More specifically, Black and White participants viewed either Black, White, or Asian targets that varied in the extent to which each violated category-based expectancies. Asian targets were included to create a condition in which both groups of participants evaluated the same out-group target.

Because the designs of Experiments 1 and 2 are different, we state the specific hypotheses and design for each separately, prior to the Method section of each. In doing so, we present our main predictions in terms of expectancy-violation theory, but for Experiment 1, we also note auxiliary predictions in terms of the complexity-extremity hypothesis and the black-sheep hypothesis.

EXPERIMENT 1

Overview and Design

We adapted Jackson et al.'s (1993) stimulus materials, which varied college application quality as well as the ethnic group of the applicant, for ease of display on computer screens. One third of the participants were randomly assigned to evaluate the Black target, one third the Asian target, and one third the White target (all targets were men). Half of the participants in each of these conditions were randomly assigned to view the applicant with strong credentials, whereas the other half viewed the applicant with weak credentials. Both Black and White participants were used. These factors constituted a 2 (strength of credentials: strong, weak) × 3 (ethnicity of target: Black, Asian, White) × 2 (ethnicity of participant: Black, White) between-subject design. Participants made affect-related evaluations and attributions by pressing a number key on the computer keyboard, which also allowed for recording each response time.

Hypotheses

In line with expectancy-violation theory, we predicted that targets who violated category-based expectations would induce extreme affect-related evaluations and that these extreme evaluations would be in the direction of the target's valence. Therefore, we predicted a main effect of target group membership on affect-related evaluations. More specifically, and based on our pretest data which confirmed that our targets violated or were consistent with category-based expectancies, we predicted that the Black target with the strong credentials would be rated more positively than either the Asian or White target. Similarly, we expected that the Asian and White targets with the weak credentials would be rated more negatively than the Black target. These effects were expected regardless of whether the perceiver was Black or White.

Next, many studies that reveal effects of category-based expectancy violations show that the valence of the target's individuating information also affects evaluations (e.g., Bettencourt et al., 1997; Biernat et al., 1996; Branscombe et al., 1993; Coleman et al., 1995; Jackson et al., 1993; Jussim et al., 1987; Jussim et al., 1996). Therefore, a main effect for the valence of the application quality was predicted such that the applicants with strong credentials would be evaluated more favorably than applicants with weak credentials (Jackson et al., 1993; Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, & Ortiz, 1982). We also included a measure of social desirability because several authors (Jackson et al., 1993; Jussim et al., 1987) have noted that White participants' evaluations of ethnic minority targets may be influenced by a desire to appear nondiscriminatory. Therefore, in an analysis of the affect-related evaluations, we were able to use this measure as a covariate to control for social desirability concerns.

We adopted Jackson et al.'s (1993) predictions for the attribution items. As such, we predicted that participants' ability and effort attributions for the targets would be affected by the group membership of the target. In other words, category-consistent information should be more strongly attributed to ability, whereas violating information should be more strongly attributed to effort. Finally, because we expected that affect-related evaluations would be more formative in the process, we predicted that partici-

pants would answer the affect-related evaluations more quickly than the causal attributions, as measured by time to press a computer key. In addition, we predicted that affect-related evaluations would mediate the effects of target group membership on causal attributions.

Although expectancy-violation theory, the black-sheep hypothesis, and the complexity-extremity hypothesis overlap in some of their specific predictions, the following are auxiliary predictions specific to the black-sheep and complexity-extremity hypotheses. The black-sheep hypothesis predicts that Black participants should evaluate the in-group applicant (Black target) with favorable credentials more positively than either the favorable Asian or White out-group applicants, but the Black participants should evaluate the in-group applicant with unfavorable credentials more negatively than similar Asian out-group and White out-group applicants. In a similar vein, White participants should evaluate in-group applicants (White target) both more positively and more negatively than either the Black out-group applicant or the Asian out-group applicant, depending on the valence of the information. In contrast, the complexity-extremity hypothesis predicts that Black participants will evaluate both the Asian applicants and the White applicants more extremely than the Black in-group applicants (i.e., evaluations of out-group members should be more polarized than those of in-group members). Complexity-extremity also predicts that White participants will evaluate both the Black and Asian applicants more extremely than the White applicants.

Method

Participants. Participants were 443 (253 women, 190 men) college undergraduates (77 Blacks, 344 Whites, and 22 others who reported no racial group) who participated for partial credit in their introductory psychology class. Due to the fact that one purpose of this study was to compare the responses of Black and White participants, the data for all other participants were dropped from the analyses, resulting in a sample size of 421.

Apparatus. Three Macintosh Quadra 700 computers and three Macintosh Classic II computers were used to present stimuli and record participant responses; all were equipped with System Software Version 7. Software used to display the stimulus materials and record participant responses was SuperLab Version 2 (Cedrus Corporation, 1991). Each computer was surrounded by a cubicle that allowed participants to hear and see the experimenter, but provided privacy while viewing the stimulus materials and responding to the dependent measures.

Materials. The target applications consisted of a photo as well as a short application form to the University of Missouri at Columbia. The application form contained the following information: a first name, birth date (similar for all applicants and intended to make it appear that all applicants were about 18 years old), intended major (math for all target applications, business or math for each filler applicant), American College Test (ACT) score, overall high-school grade point average (GPA), racial or ethnic group, and gender.

Ethnicity of the target (Black, Asian, or White) was manipulated by a check mark on the application and a target photo. The three photos were chosen from a set of 50 on the basis of a pretest. On 7-point scales ranging from 1 (*not at all*) to 7 (*very much*), participants (Black, Asian, and White students: $N = 50$) showed similar levels of rated attractiveness (Asian: $M = 2.76$, Black: $M = 2.79$, White: $M = 2.88$), likability (Asian: $M = 2.35$, Black: $M = 2.70$, White: $M = 2.76$), intelligence (Asian: $M = 2.20$, Black: $M = 2.88$, White: $M = 2.44$), and studiousness (Asian: $M = 3.36$, Black: $M = 2.36$, White: $M = 3.40$). In addition, half the applications indicated a high GPA (3.80) and ACT score (95%), whereas the other half indicated a low GPA (2.02) and ACT score (75%). The filler applications were intended to be of average quality and indicated GPA scores of 3.01 or 3.00 and ACT scores of 88% and 92%. All identifying information (e.g., last names, signatures, and social security numbers) appeared to have been blacked out of the application form. In a pretest of the materials, participants ($N = 212$; pretest participants were Black, Asian, and White) rated how much each target fit the stereotype for the respective ethnic group on 5-point scales ranging from 1 (*does not fit the stereotype for Blacks, Asians, Whites)* to 5 (*fits the stereotype for Blacks, Asians, Whites*). The results indicated that the skillful Black applicant was rated as fitting the respective stereotype less ($M = 1.70$) than either the skillful Asian ($M = 4.12$), $t(70) = -6.14$, $p < .05$ ($d = 1.45$), or the skillful White applicant ($M = 3.56$), $t(69) = -4.68$, $p < .01$ ($d = 1.11$). In contrast, the unskillful Asian ($M = 1.78$), $t(70) = -6.10$, $p < .05$ ($d = 1.44$) and White applicants ($M = 2.69$), $t(70) = -2.31$, $p < .05$ ($d = 0.55$), were rated as fitting the respective stereotypes less than the unskillful Black applicant ($M = 4.20$). These results did not differ by participant ethnic group.

Dependent measures. To assess participants' impressions of the applicants, we used measures adapted from Jackson et al. (1993) as well as Bettencourt et al. (1997).[1] Participants first answered the evaluation questions, indicating

[1] After participants made attributions and evaluations of each target, they were asked to answer a series of other questions and assessments that we had thought may mediate (e.g., cognitive complexity) or moderate (e.g., in-group identity) the evaluations of the targets. However, the position of these measures was less than ideal. Taking these measures at the end of the study makes it impossible to determine whether the process of making evaluations and attributions influences, in some way, the responses to these latter measures. Therefore, the results for these measures are not discussed.

how favorable, good, and likable each applicant was (e.g., "After reading the application, do you think that Daniel would be considered a good or bad applicant?") on 9-point scales ranging from 1 (*very bad, unfavorable, unlikable*) to 9 (*very good, favorable, likable*). Next, participants answered three attribution questions also used by Jackson et al. (1993). They assessed the importance of luck, ability, and effort in explaining the quality of the applicant's credentials. Higher values indicated that the characteristic was more important in explaining the applicant's performance than lower numbers (e.g., "Does Daniel's ability explain the quality of his application?") on 9-point scales ranging from 1 (*not at all explains*) to 9 (*very much explains*). Several additional attribution questions (also adapted from Jackson et al., 1993) assessed the importance of academic assistance programs, support from teachers, and the grading standards of his high school in explaining the applicant's performance. Again, 9-point scales ranging from 1 (*not at all explains*) to 9 (*very much explains*) were used with higher values indicating that the characteristic was more important in explaining the applicant's performance. The computer randomly ordered all questions within type (evaluations, attributions) before they were shown to the participants. Finally, because we were interested in assessing reaction time to the attribution and evaluation questions, we conducted a pretest to determine the read times associated with our measures. This pretest ($N = 32$) showed that the read times for all questions were fairly equivalent (read times in seconds were: ability: $M = 35$, $SD = 11$; effort: $M = 33$, $SD = 11$; task attributions: $M = 37$, $SD = 11$; affect-related evaluations: $M = 38$, $SD = 8$), ensuring that any differences in response time would not be a result of differential read time.

Procedure. Participants first were given an informed consent form followed by a short demographic form that asked for their gender, racial group, age, and year in college. Participants then were seated at one of seven randomly assigned computers while the experimenter read the instructions and explained the purpose of the study. There were 3 to 7 participants in each session.

Participants were told that several universities were considering receiving undergraduate applications via the Internet and, thus, the purpose of the study was to investigate how these types of applications (those examined via computer) may be judged in comparison to the conventional paper-and-pencil applications. Participants were further informed that sometimes applicants send pictures of themselves with their applications, which also could be sent via the Internet. Finally, participants were told that they would see information regarding several college applicants and that they would answer questions about each.

After hearing these instructions, participants began reading through each screen of the computer program. The first few screens consisted of additional instructions designed to help the participants in using the computers and to familiarize them with pressing the number keys. Next, the applicant's photo was shown first (shown for 15 sec) followed by the application form (shown for 33 sec). Having examined the application, participants answered the dependent measures for that applicant. The computer program permitted participants to continue to each subsequent screen of the dependent measures at their own pace by answering the question with a number key.

Following the completion of the dependent measures, participants answered several manipulation check questions regarding each applicant's gender, major, and GPA. The Marlowe–Crowne social desirability scale was administered at the conclusion of the study (see Fischer & Fick, 1993), in addition to a suspicion check questionnaire. Participants were debriefed by mail.

Results

Six participants with obvious response sets (answering with an identical response or unrealistically short response time for all questions) and 5 participants who failed to complete all of the tasks of the experiment (most of whom were also suspicious) were dropped from the analysis (4 Blacks, 7 Whites). Therefore, analyses were completed with the data from 410 participants. Due to the fact that the reliability of the evaluation index was adequate ($\alpha = .93$), the three items were averaged. In what follows, we report four series of analyses:

1. An examination of the residualized data.
2. Analyses of variance for affect-related evaluations and causal attributions.
3. Within-subjects comparisons of response times.
4. Regression analyses designed to test mediation.

Prior to analyses, all data were examined for violations of normality and extreme outliers.[2]

Bivariate relations. To examine the correlations among evaluations and attributions with the experimental effects of target race and application quality removed, a set of residual scores were calculated, followed by a correlational

[2]According to the recommendation of Tabachnick and Fidell (1989), Tukey (1990), and Wilcox (1995), outliers were identified examining box plots of the data. A box plot is a graphical display that indicates the range, the quartiles, the spread, and the median of the data. The length of the box plot represents 50% of the cases, and thus the lower boundary of the box is the 25th percentile and the upper boundary is the 75th percentile. Data points with values more than three box lengths from the upper or low boundary of the box are extreme outliers (Tabachnick & Fidell, 1989; Wilcox, 1995). Three data points of a total of 2,475 were extreme outliers (less than 1% of the data). As recommended by Tabachnick and Fidell and Wilcox, these extreme outliers were modified to the value of the next closest nonoutlying value in the distribution.

analysis of these scores. As can be seen in Table 1, the correlations among the attribution ratings and the affect-related evaluations are smaller than those found among the attribution ratings.

Affect-related evaluations. Table 2 presents the mean affect-related evaluations for the three targets. As shown in the table, the Black applicant with the strong credentials was evaluated more positively than the Asian and White applicants with similar credentials, and the Asian and White applicants with weak credentials were evaluated more negatively than the Black applicant with weak credentials. The predicted main effect of target's ethnic group membership was reliable, $F(2, 404) = 6.27, p < .01$ ($d = 0.32$). Although differences in the evaluations are much weaker for the Black participants, the patterns are very similar across participant groups. Moreover, the analysis revealed no interactions with participants' ethnicity (three-way interaction), $F(2, 398) = 0.09, p > .90$ ($d = 0.02$). Planned comparisons combining Black and White participants' data show that the Black applicant with the strong credentials was evaluated more positively than either the Asian, $t(122) = 1.40, p < .10$ ($d = 0.25$), or the White, $t(131) = 1.96, p < .05$ ($d = 0.34$), applicant with the same credentials. However, the Asian and White strong credential applicants were rated similarly, $t(132) = 0.57, p > .25$ ($d = 0.09$). In the weak credential conditions, the Black applicants were evaluated more positively than either the Asian, $t(139) = 2.82$, $p < .01$ ($d = 0.48$), or the White applicants, $t(140) = 2.76, p < .01$ ($d = 0.46$). Just as in the strong credential conditions, there was no significant difference between the Asian and the White applicants, $t(150) = 0.06, p > .25$ ($d = 0.01$). Finally, applicants with strong credentials were evaluated more positively than those with weak credentials, $F(1, 404) = 1319.66$, $p < .01$ ($d = 3.59$).

When scores on the Marlowe–Crowne social desirability scale were used as covariates (as a control for social desirability concerns), the results were essentially unchanged: main effect for target's ethnic group membership, $F(2, 387) = 5.54$, $p < .01$ ($d = 0.17$), and main effect for application quality, $F(1, 387) = 1259.36, p < .01$ ($d = 2.54$).

TABLE 1
Bivariate Associations Between Evaluations and Attributions: Controlling for Effects of Applicant Race and Application Quality

	Evaluations	*Ability Attribution*	*Effort Attribution*
Ability attribution	.25*		
Effort attribution	.13*	.65*	
Task attribution	.17*	.41*	.41*

*$p < .01$.

TABLE 2
Mean Values and Standard Deviations for Affect-Related Evaluations

	Targets					
	Strong Credentials			*Weak Credentials*		
Participants	*Asian*	*Black*	*White*	*Asian*	*Black*	*White*
Black						
M	8.08	8.13	8.07	4.33	4.74	4.67
SD	1.56	0.76	0.39	1.40	1.01	1.28
White						
M	8.16	8.42	8.03	4.36	4.87	4.31
SD	0.77	0.58	0.73	1.36	1.20	1.10
All participants						
M	8.14_{ab}	8.38_{a}	8.04_{b}	4.36_{c}	4.84_{d}	4.37_{c}
SD	0.85	0.60	0.68	1.36	1.16	1.13

Note. For Black participants, cell frequencies ranged from 10–14. For White participants, cell frequencies ranged from 50–64. For all participants, cell frequencies ranged from 64–83. Higher values indicate more favorable affect-related evaluations of the target. Row means that do not share a subscript differ reliably, denoting at least $p < .05$ (planned t tests).

Speed of evaluations and attributions. Prior to the analyses, the distributions of response times were examined for values that were either extremely short or long compared to the rest of the data. Reaction times less than 1,000 msec and those over 14,000 msec were identified as outliers, and were not included in the analyses (Lambert, 1995; Newman, 1993; Smith & Miller, 1979). Note that this upper cutoff, although unusually long, is likely a reflection of the fact that participants were not instructed to answer quickly, as often they are in response time studies. Distributions of response times showed no evidence of positive skew following the application of this cutoff.

The response times were unaffected by the race of the targets and by the interaction between target race and application quality (Fs = .15–2.05, ps > .12, ds < .10). However, following weak applications, participants responded less quickly to evaluations, $F(1, 54) = 5.20, p < .05$ ($d = 0.44$; $M = 5{,}040.30$, $SD = 1{,}259$; $M = 5{,}743.33$, $SD = 1{,}001$ for strong and weak applications, respectively), ability attributions, $F(1, 53) = 4.53, p < .05$ ($d = 0.41$; $M = 5{,}412.44$, $SD = 2{,}825$; $M = 6{,}725.89$, $SD = 2{,}057$ for strong and weak applications, respectively) and effort attributions, $F(1, 49) = 13.55, p < .01$ ($d = 0.74$; $M = 4{,}255.00$, $SD = 1{,}480$; $M = 6{,}802.20$, $SD = 3{,}086$ for strong and weak applications, respectively).[3] Speed of responses to task-related attributions was not significantly affected by applicant quality, $F(1, 54) = 1.85, p > .15$ ($d =$

[3]Other research (Hamilton & Zanna, 1972) has shown that people seem to pay more attention to negative information.

0.26; (M = 5,375.14, SD = 2,067; M = 6,133.23, SD = 1,823 for strong and weak applications, respectively). The results suggest that responses to the affect-related evaluations were quicker than those to the causal attributions. As shown in Table 3, affect-related evaluations were made more quickly than effort attributions, $F(1, 393) = 5.92$, $p < .01$ ($d = 0.12$), task attributions, $F(1, 406) = 19.10$, $p < .01$ ($d = 0.22$), and ability attributions, $F(1, 387) = 33.51$, $p < .01$ ($d = 0.29$).

Causal attributions. As predicted, the pattern of attributions was similar to that found by Jackson et al. (1993). The analysis of the effort attributions revealed a main effect of targets' ethnic group, $F(2, 404) = 3.33$, $p < .05$ ($d = 0.23$). In the strong credential conditions, effort attribution ratings were higher when the target was Black (M = 7.00, SD = 2.18) than when the target was Asian (M = 6.34, SD = 2.56) or White (M = 6.94, SD = 1.87). However this difference was only significant for the Black versus Asian comparison, $t(122) = 1.70$, $p < .05$ (d = 0.30). Among targets with weak credentials, the Asian (M = 4.72, SD = 2.55) target was assigned lower effort attribution ratings than the Black (M = 5.38, SD = 1.99) target, $t(139) = 1.84$, $p < .05$ ($d = 0.31$). The Asian target with weak credentials was also rated lower on effort than the similar White (M = 5.23, SD = 2.32) target, although this difference was only marginally significant, $t(150) = 1.40$, $p < .10$ (d = 0.23).

The task-related attributions were also affected by the target's ethnic group, $F(2, 404) = 6.24$, $p < .01$ ($d = 0.35$). The means tended to show that for both the targets with strong credentials and the targets with weak credentials, the White target was given the highest task attribution ratings. Among targets with strong credentials, the quality of the White (M = 5.10, SD = 2.08) target's application was attributed more to task-related variables (e.g., grading standards) than was the quality of the Asian (M = 4.59, SD = 2.14) target's application, $t(132) = 2.32$, $p < .01$ ($d = 0.40$) or the Black (M = 4.61, SD = 1.87) target's application, $t(131) = 2.28$, $p < .05$ (d = 0.39). Among targets with weak credentials, the White (M = 4.45, SD = 1.95) target was given the highest task-related attribution ratings as compared to the Black (M = 3.72, SD = 1.84) target, $t(140) = 2.70$, $p < .01$ ($d = 0.45$) and the Asian (M = 3.43, SD = 1.65) target, $t(150) = 3.19$, $p < .01$ ($d = 0.52$).

Finally there was no effect of targets' ethnic group on the ability attributions, $F(2, 404) = 1.45$, $p > .20$ ($d = 0.12$). No differences were found among the strong credential applicants (Ms = 6.96, 7.23, and 7.07; SDs = 2.38, 2.07, and 1.99, for Asian, Black, and White applicants, respectively). Among the weak credential applicants, the Asian (M = 4.46, SD = 2.31) target was rated lower than either the Black (M = 5.12, SD = 2.37) target, $t(139) = 2.09$, $p < .05$ ($d = 0.35$) or the White (M = 5.02, SD = 2.43) target, $t(150) = 2.03$, $p < .05$ (d = 0.33).

Overall, the attributions tended to follow the pattern shown by Jackson et al. (1993); that is, violating behaviors were attributed more to unstable and external factors (effort), whereas consistent behaviors were attributed to more stable factors (ability, task).

TABLE 3
Mean Response Times to Evaluations, Effort, Task, and Ability Attributions in Experiment 1

Variable	*M*	*SD*
Affect-related evaluations	$5{,}086_a$	1,393
Effort attribution	$5{,}400_b$	2,923
Task attribution	$5{,}535_b$	2,131
Ability attribution	$5{,}805_c$	2,695

Note. Response times are recorded in msec; higher numbers indicate longer times to respond to the questions. Mean response times with unique subscripts differ, denoting at least $p < .05$ (planned t tests).

Regression analyses. We conducted a series of regression analyses as outlined by Baron and Kenny (1986) to test whether affect-related evaluations mediate the relation between target ethnicity and causal attributions. To meet the first criterion for mediation, the results must show that the independent variable (target ethnicity) affects the mediator (evaluations). This criteria was met because, as previously reported, the analyses revealed a main effect of target ethnic group membership on the affect-related evaluations. Second, the independent variable must affect the outcome variable (attributions). The previous analysis also revealed main effects of target ethnic group membership on the effort and task attributions. Finally, to satisfy the third criterion for mediation, the effect of the independent variable on the outcome variable should be rendered nonsignificant when entered into simultaneous regression with the mediating variable. As shown in Table 4, when affect-related evaluations and the independent variable (i.e., target ethnic group) were simultaneously entered as predictors of the effort and ability attributions, the effect of the independent variable was nonsignificant. The results of these analyses suggest that affect-related evaluations may mediate attributional processes.

Tests of auxiliary hypotheses for evaluations. The complexity-extremity hypothesis predicted that evaluations of out-group members should be more polarized whereas the black-sheep hypothesis predicts that evaluations of in-group members should be more polarized. Our results, however, showed no evidence for either. Interactions between target valence and target ethnicity were nonsignificant for White, $F(2, 333) = 1.83$, $p > .10$ ($d = 0.10$) and Black participants, $F(2, 70) = 1.28$, $p > .10$ ($d = 0.18$). The complexity-extremity hypothesis predicts more range in evaluations directed at out-group targets than those directed at in-group targets. However, the Black participants evaluated the Black targets

TABLE 4
Regression Analyses to Test Affect-Related Evaluations as Mediators of Attributions in Experiment 1

Predictors	β	*t Value*
Effort attribution		
Target ethnicity	.04	0.89
Affect-related evaluation	.35	7.61*
Task attribution		
Target ethnicity	−.08	−1.66
Affect-related evaluation	.28	5.91*

*$p < .001$ or smaller (planned t tests).

and the White targets with similar range. For the Black targets, the evaluations ranged from 8.13 (SD = .71) for the strong credential target to 4.74 (SD = 1.01) for the weak credential target (range = 3.39). A similar range was obtained for the White targets (strong credential target: M = 8.07, SD = .39; weak credential target: M = 4.67, SD = 1.27; range = 3.40). However, the range for the Asian targets was considerably larger (strong credential target: M = 8.08, SD = 1.56; weak credential target: M = 4.33, SD = 1.41; range = 3.75). In contrast, White participants evaluated the Asian targets and White targets with similar range. For the White targets, the evaluations ranged from 8.03 (SD = .74) for the strong credential target to 4.31 (SD = 1.10) for the weak credential target (range = 3.72). A similar range was obtained for the Asian targets (strong credential target: M = 8.24, SD = .58; weak credential target: M = 4.87, SD = 1.20; range = 3.80). However, the range for the Black targets was smaller (strong credential target: M = 8.16, SD = .77; weak credential target: M = 4.36, SD = 1.36; range = 3.55).[4]

Discussion

Overall, the results of Experiment 1 were consistent with the predictions of the expectancy-violation theory. First, supporting our proposed addition to Jackson et al.'s (1993) model, the results showed that affect-related evaluations were made more quickly than ability, effort, and task attributions. In addition, target ethnic group membership affected the responses to the effort and the task attribution questions, allowing us to conduct regression analyses to examine whether the relation between target group membership and causal attributions was mediated by affect-related evaluations. The results of this mediation analysis provide strong initial support for our prediction that participants' affect-related evaluations may be related to their causal attributions. Indeed these results suggest complete mediation (e.g., Baron & Kenny, 1986) of the relation between target group membership and causal attributions. These results suggest that initial affect-related evaluations account for much of the variability in these subsequent attributional judgments, and perhaps other assessments as well (e.g., predictions for future success).

Second, the results for the affect-related evaluations showed similar effects of category-based violations among both the Black and the White participants. This finding supports our hypothesis that the effects of expectancy violation are similar across ethnic groups. As such, within one study, these results provide more definitive evidence for the continuity in responses across evaluators than was suggested in the separate studies of Jussim et al. (1987) and Coleman et al. (1995). Recall that the pattern of results found in Coleman et al. for Black participants was essentially identical to that for the pattern found among White participants by Jussim et al.

Although the affect-related evaluations failed to yield results that strongly supported either the black-sheep or the complexity-extremity hypothesis, some of the patterns were consistent with their predictions. For example, supporting the complexity-extremity hypothesis, the results revealed that for both the White and Black participants, evaluations of the Asian targets had the most range (perhaps demonstrating a less complex cognitive schema for this out-group target). However, among the Black participants, the range of the evaluations for the White targets was very similar to that for in-group targets. This finding is somewhat consistent with Coleman et al.'s (1995) finding that Black participants do not evaluate White targets with more range than Black targets. As Coleman et al. suggested, it may be that, similar to White participants, ethnic minority participants (or, perhaps, smaller or stigmatized groups) tend to show relatively restricted range in their evaluations of White "majority" targets. This finding may be the result of less restrictive stereotypes for majority targets, or perhaps because Black perceivers (especially college students) are frequently exposed to Whites and White culture.

Several other areas of research provide evidence bearing on this issue. For example, work in stereotyping (Fiske, 1993) suggests that those who have less power (e.g., disadvantaged minority groups) may be the least likely to use stereotypes and may be the most motivated to know and understand those who have greater power. In addition, other theorists (e.g., Roediger, 1998) have suggested that Blacks may be especially knowledgeable about Whites as White culture often defines many of the norms and standards of American life (see also McIntosh, 1992).

[4]Of lesser importance to this work, at the end of Experiment 1 we also measured cognitive complexity (Linville & Jones, 1980). Similar to the findings of other research (Coleman et al., 1995), the results showed that participants' schemas were more complex for their in-groups than their out-groups. Black participants' scores suggested more cognitive complexity for Black male undergraduates (M = 4.42) than for either Asian or White male undergraduates (M = 4.19 and 3.88, respectively), and White participants' scores suggested more cognitive complexity for White male undergraduates (M = 4.95) than Asian or Black male undergraduates (M = 3.74 and 4.24, respectively).

EXPERIMENT 2

As we have noted, consistent with our modification to Jackson et al.'s (1993) process model, the findings of Experiment 1 show that affect-related evaluations were made more quickly than causal attributions. However, as noted by Fazio (1990), although numbered rating scale measures often provide more complex information regarding participants' responses, the use of a response format involving multiple response options may artificially increase the variability of observed latencies. Therefore, in Experiment 2 we changed the paradigm so that we could test the model using a forced choice, yes–no response format. This response paradigm also allowed us another advantage. Although our regression analyses in Experiment 1 provided some initial support for the idea that evaluations mediate attributions, the response paradigm that we adopted for Experiment 2 allowed us to calculate facilitation scores designed to further test mediation (Bassili & Racine, 1990).

In addition to these differences, in Experiment 2 we were able to remedy another shortcoming of Experiment 1. The order of attribution and affect-related evaluation questions was counterbalanced in Experiment 2, thereby allowing us to draw firm conclusions about the expected facilitation effects.

Due to the fact that we felt that Experiment 1 sufficiently supported our hypotheses for the secondary purpose of our research (examining category-based expectancy violation with more than one participant group membership), this question was not addressed in Experiment 2. In summary, Experiment 2 focuses exclusively on determining the sequence of events that unfold on exposure to category-based expectancy violation.

Overview and Design

The method we used in Experiment 2 was based on Bassili and Racine's (1990) judgment facilitation paradigm. This paradigm allowed us to use a greater variety of targets in a completely within-subjects design. Participants were presented with sentences describing group members whose actions were either consistent with or in violation of the stereotype for their group. Each sentence was followed by two of three potential pairs of questions concerning the cause of the person's behavior described in the sentence (person or situation), an evaluation of the person (good or negative), or control questions (e.g., Bassili & Racine, 1990; Smith & Miller, 1983). For ease of explanation, the attribution and evaluation questions are collectively referred to as criterion questions. The possible pairs of questions were control–attribution, control–evaluation, evaluation–attribution, and attribution–evaluation. This methodology allowed us to assess whether the response to the first question in a pair facilitates (i.e., speeds up) the response to the second question. The order of questions was counterbalanced among the violating and consistent sentences, such that all possible combinations appeared equally with both types of targets. Thus, the design was a 2 (targets; consistent, violating) × 3 (question type; attribution, evaluation, control) × 3 (question order; attribution first, evaluation first, control first) completely within-subjects factorial design.

Hypotheses

The logic of this paradigm has two critical components for this research. First, if in the process of answering a first question, information relevant to the second question is activated, the time required to answer the second question should be reduced (i.e., facilitated) compared to the case in which the latter information is not activated (Bassili & Racine, 1990; Klein, Loftus, & Burton, 1989). If, as we have argued, affect-related evaluations occur prior to causal attributions in the process, then evaluation questions that follow attribution questions should be answered more quickly than evaluation questions that follow control questions (i.e., having to make an attribution should facilitate the time needed to make an evaluation). This is because the evaluative information should already be activated in the former case. Second, for a given variable (i.e., evaluation or attribution) to be considered a mediator of another variable, the proposed mediating variable must be responded to more quickly than the variable it mediates (Bassili & Racine, 1990; Smith & Miller, 1983). In accordance with this logic, our model predicts that affect-related evaluations should be made more quickly than attributions.

A final prediction concerns the expectedness of target behavior. According to Greenberg and Pyszczynski (1982), attributions for expectancy-consistent information should be ready made, and as such should be more accessible following expectancy confirmation. However, when an expectancy is violated, these stored attributions are insufficient for explaining the target's actions, and thus should be supplanted by other attributions. No such assumption is made regarding evaluations. Based on these theorist's predictions, we hypothesized an interaction between type of judgment (evaluation vs. attribution) and target condition (consistent vs. violating), such that, whereas level of expectancy violation should not affect the speed of evaluations, attributions to violating targets would be slower than attributions to consistent targets.

Method

Participants. Participants were 89 undergraduates (44 women, 45 men) who received partial course credit in their introductory psychology class.

Materials. Target sentences consisted of information concerning the target's category membership, their name, and a behavior that was either consistent with or violated cate-

gory-based expectancies for their group. Examples of expectancy-consistent target sentences include: "The environmentalist, Ronnie, helped clean up the river," and "The feminist, Arlene, yelled in the man's face." Examples of expectancy-violating target sentences include: "The welfare mother, Patrice, graduated from medical school," and "The Asian student, Tim, earned poor grades in Math." The 32 target sentences (half included a category-based expectancy violation) were chosen from a total of 94 on the basis of pretest ratings. The pretest questionnaire asked participants to rate the extent to which each sentence fit with the common stereotype for a given group (e.g., vegetarian, environmentalist) using a 6-point rating scale ranging from 1 (*extremely consistent with the group stereotype*) to 6 (*extreme violation of the group stereotype*). The mean response for the consistent sentences was 1.65 (SD = .38) and 5.09 for the violation sentences (SD = .52).

Procedure. After completion of informed consent forms and listening to preliminary instructions, participants were randomly assigned to one of six computer terminals. The experimenter then read some additional instructions concerning the use of the computer keyboard in navigating through the task, after which the majority of the instructions for the task were presented via computer.

The computerized instructions informed participants that they would be reading sentences that described a person and that person's action, each followed by two questions that were to be answered by pressing a "yes" or "no" key on the keyboard (Z and M keys marked with a yellow *Y* and *N*). The keys assigned to "yes" and "no" responses were counterbalanced across sessions, such that "yes" appeared on the right side in half the sessions.

As in similar paradigms (Bassili & Racine, 1990; Smith & Miller, 1983), participants were trained to recognize specific questions via single probe words representing those questions. For example, participants were taught to consider the question "Did something about the situation cause the behavior described in the sentence?" when they were presented with the single probe *situation.* Other, similar probe words were established to represent questions concerning person attributions, affect-related evaluations, and control questions (e.g., "Was the question printed on a single line of the computer screen?"). Using the single-word probes prevents confounding question reading times with response times. The training task began with the presentation of the probe words along with their definitions. Each probe was then presented alone and participants were instructed to press a response key as soon as they could think of its full meaning. The full definition of the probe appeared following the key press. Next, participants were presented with eight practice sentences similar to those used in the remainder of the experiment, and responded to two questions (with the word probe procedure) following each one. Participants were given several opportunities to ask for clarification prior to and following the practice trials.

During the experimental trials, when a target sentence appeared on the screen, it remained until participants pressed the space bar to indicate they had read it. Immediately afterward, the first of two question probes appeared on the screen and remained until the participant answered by pressing either the "yes" or "no" key. The second question probe then followed.

After the computer task, participants answered a short measure designed to reveal suspicion of the hypothesis and then were debriefed and dismissed (no participant revealed awareness of the hypothesis).

Results

As did Bassili and Racine (1990), we used the median of each participant's response times for each of the target conditions. Using medians eliminates the influence of extreme responses and reduces the positive skew in the data (Fazio, 1990). As such, analyses were based on the mean of participants' median responses in each target condition. We calculated evaluation facilitation scores by subtracting the median response time to make an evaluation after it followed an attribution probe from the median response time to make an evaluation after it followed a control probe. Likewise, we calculated attribution facilitation scores by subtracting the median time to make an attribution that followed an evaluation probe from the median time to make an attribution that followed a control probe. A positive facilitation score indicates that answering the first criterion question facilitates (i.e., speeds up) answering the second criterion question. In the following sections, we present analyses for the facilitation scores as well as for the raw response latencies.

Bivariate relations. To assess redundancy in our dependent measures, we calculated bivariate correlations among our raw latency measures and facilitation scores for both attributions and affect-related evaluations separately for consistent and violation conditions. These correlations are presented in Table 5. As the table shows, raw latency and facilitation score measures of attributions and affect-related evaluations were similar within experimental conditions, with correlations ranging from .70 to .86, and were much smaller across experimental conditions.

Facilitation scores. The facilitation scores for evaluations and attributions are presented in Table 6. Facilitation scores were analyzed using a series of paired-sample t tests to determine whether the scores differed from zero. The results showed that evaluations were made more quickly when they followed an attribution than when they followed a control

TABLE 5
Bivariate Correlations Among All Measured Variables By Experimental Condition: Experiment 2

	1	2	3	4	5	6	7
Expectancy Consistent							
Attribution (F)	—						
Evaluation (F)	.03	—					
Attribution (R)	.86*	.02	—				
Evaluation (R)	−.05	.84*	−.11	—			
Expectancy Violation							
Attribution (F)	−.14	−.22**	−.03	−.28*	—		
Evaluation (F)	−.11	.21**	−.03	.17***	−.04	—	
Attribution (R)	−.19***	−.08	−.05	−.14	.83*	−.02	—
Evaluation (R)	−.09	.10	−.03	.09	.08	.70*	.06

Note. F = facilitation score (see text for description of calculation procedures); R = raw median response latency.

TABLE 6
Mean Facilitation Scores and Standard Deviations as a Function of Target Condition and Judgment Type: Experiment 2

	Target Condition			
	Consistent		*Violation*	
Judgment Type	*M*	*SD*	*M*	*SD*
Evaluations	371.12*	528.60	360.00*	405.58
Attributions	32.10	861.23	136.18	1169.85

Note. Scores and standard deviations expressed in msec. See text for an explanation of how facilitation scores were calculated.
*$p < .01$.

probe (i.e., facilitation scores for the evaluation probes were significantly greater than zero) in both the consistent target condition, $t(88) = 6.62, p < .001$ ($d = 0.69$), and the violating target condition, $t(88) = 8.35, p < .001$ ($d = 0.88$). In contrast, the difference between attributions made following evaluations and attributions made following control questions revealed no facilitation in either the violating, $t(89) = 1.09, p > .25$ ($d = 0.11$), or consistent target conditions, $t(89) = 0.35, p > .50$ ($d = 0.03$). Therefore, these results indicate that whereas making an attribution in the first position facilitated evaluations made in the second position, making evaluations in the first position did not facilitate attributions in the second position. This pattern is consistent with the hypothesis that evaluations mediate attributions because facilitation suggests that making attributions in the first position necessitates activation of target evaluations.

Raw latencies. Table 7 presents the average response latencies (in msec) for evaluations and attributions by target condition, which were analyzed with a 2 (judgment type: evaluation, attribution) × 2 (target condition: violation, consistent) × 2 (question order: control first, criterion first) analysis of variance. The analysis revealed a significant main effect of judgment type, $F(1, 88) = 132.74, p < .001$ ($d = 1.21$), indicating that across target conditions, evaluations (M = 1,276.32 msec, SD = 266.26) were made significantly more quickly than were attributions (M = 2,100.85 msec, SD = 783.08). This finding is consistent with the pattern of facilitation scores reported previously. Of greater interest was the significant interaction between judgment type and target condition, $F(1, 88) = 5.30, p < .05$ ($d = 0.24$). Inspection of the response latencies shows that whereas response times to make evaluations were not significantly affected by target condition, the time to make attributions about expectancy-violating targets was significantly slower than that required to make attributions following expectancy-consistent targets. In addition, significant main effects of target condition, $F(1, 88) = 7.39, p < .01$ ($d = 0.29$), and question order, $F(1, 88) = 35.79, p < .01$ ($d = 0.63$), indicated that responses to expectancy-violating targets were made more slowly (M = 1,753.40 msec, SD = 569.77) than those to consistent targets (M = 1,623.78 msec, SD = 482.39), and that responses to second questions were quicker when the first question was a criterion (M = 1,576.28 msec, SD = 481.57) than when it was a control (M = 1,800.89 msec, SD = 535.76).

Discussion

The results of Experiment 2 support the findings of Experiment 1 showing that affect-related evaluations are made more readily than attributions. Moreover, the facilitation scores provided further evidence that evaluations mediate attributions. In addition, these results support prior research showing that affective evaluations may be made relatively automatically (Bargh et al., 1992), whereas attributions require more effortful processing (Smith & Miller, 1979).

TABLE 7
Mean Response Latencies and Standard Deviations as a Function of Target Condition and Judgment Type: Experiment 2

	Target Condition			
	Consistent		*Violation*	
Judgment Type	*M*	*SD*	*M*	*SD*
Evaluations	1,260.99$_a$	353.42	1,291.66$_a$	261.23
Attributions	1,986.56$_b$	744.82	2,215.13$_c$	993.54

Note. Mean response latencies and standard deviations expressed in msec. Means that do not share a subscript differ, denoting $p < .05$ (planned t tests).

Finally, although the results generally showed that response times to make evaluations were faster than those to make attributions, the results of Experiment 2 provide initial support for the idea that attributions may be more effortful when targets violate category-based expectancies than when they are consistent with those expectations; that is, the significant interaction between judgment type and target condition showed that the time to make attributions about expectancy-violating targets was significantly slower than that to make attributions about expectancy-consistent targets.

GENERAL DISCUSSION

The primary purpose of this research was to further articulate the processes that unfold from category-based expectancy violations. Expectancy-violation theory (Jussim et al., 1987) generally predicts that persons who violate category-based expectancies will be evaluated more negatively or more positively than persons who do not, depending on the direction of the valence of target information. Jackson et al. (1993) further developed the predictions of expectancy-violation theory by proposing a preliminary process model of category-based expectancy violations. In their model, Jackson et al. specified that perceivers make causal attributions that help explain target information that violates expectancies. Moreover, they argue that these explanations then mediate evaluations of the target. We (Bettencourt et al., 1997) have argued that, on exposure to targets who violate category-based expectations, perceivers make initial evaluations associated with their immediate affective reaction to the violation. In this work, we further suggest that effortful, attributional processes are unlikely to precede this type of affect-related evaluation. To clarify the sequence of these processes, in Experiments 1 and 2, we compared the time it took participants to make evaluations of and attributions about targets that either violated or were consistent with category-based expectations. The results from these experiments provide reliable evidence supporting the prediction that initial affect-related evaluations are made more quickly than attributions. Moreover, both the regression analyses designed to test mediation in Experiment 1 and the analyses of the facilitation scores of Experiment 2 supported the prediction that affect-related evaluations mediate attributional explanations.

These findings are consistent with prior evidence suggesting a primacy of affective evaluations. For example, Bargh et al. (1992) showed that evaluations of global favorability—or good versus bad—are made relatively quickly. Moreover, our prediction that violations of expectations may induce affective reactions concurs with emotion theorists' claims that unexpectedness is an intensity variable of emotion (Clore, Schwarz, & Conway, 1994). The literature revealing effects of unexpected events on emotional response (e.g., MacDowell & Mandler, 1989) as well as Beirnat et al.'s (1996) recent data showing category-based violation effects on negative mood change support our argument that affect-related evaluations ensue directly from category-based expectancy violations. However, in future research it may be useful to include not only affective evaluations (e.g., favorable, good), but also measures of specific types of affective reaction (e.g., surprise). Such measures would allow us to further examine how affective reactions like surprise affect the sequence of processes that unfold from category-based expectancy violation. Regardless of this distinction, subsequent to emotional reactions perceivers may engage in attributional processes that in turn, as specified by Jackson et al. (1993), influence subsequent judgments about a category-violating target. As noted earlier, these later judgments may be more cognitive in nature because they follow attributional explanations.

To increase understanding of category-based expectancy effects, researchers have noted that studies should include participants who vary in terms of their ethnic group membership. To that end, a secondary purpose of our research was to include both Black and White participants. Expectancy-violation theory suggests that regardless of perceivers' in-group membership, they should evaluate category-violating targets more extremely than similar category-consistent targets. Despite this prediction, very few studies (e.g., Bettencourt et al., 1997) have included members from more than one group and none have included participants who vary by ethnic group membership. This is particularly important, because many of the studies that test the predictions of expectancy-violation theory (Bettencourt et al., 1997; Biernat et al., 1996; Jackson et al., 1993; Jussim et al., 1987; Jussim et al., 1996) have varied the ethnic group membership of the target (particularly the ethnic group categories of Black and White) but not that of the perceiver. As predicted by expectancy-violation theory, the results of Experiment 1 reveal a similar pattern of extremity for the affect-related evaluations among the different groups of participants. Regardless of their group membership, perceivers evaluated both in-group and out-group targets more extremely when the respective information violated category-based expectations.

By including participants who varied in their group membership, our experiments also addressed several other theoretical perspectives that have sought to explain evaluative extremity. As noted, the complexity-extremity hypothesis predicts that perceivers will evaluate out-group members more extremely (Linville & Jones, 1980; Marques et al., 1988), whereas the black-sheep hypothesis predicts that perceivers will evaluate in-group members more extremely. Although supportive evidence has been found for each of these predictions, studies designed to test them have failed to include perceivers of more than one group (e.g., ethnic group). These results show instances of extremity in the evaluations of both in-group and out-group targets, but our results did not reveal that in-group perceivers either

consistently evaluated the out-group targets more extremely (i.e., out-group polarization) or consistently evaluated the in-group targets more extremely (i.e., in-group polarization). If this had been the case, we would have observed stark differences between the patterns of evaluations for each of the participant ethnic groups. Instead, as noted, the patterns of evaluative extremity were similar among the two groups of participants.

This continuity in evaluative patterns is consistent with our (Bettencourt et al., 1997) argument that the apparent contradictions between Linville and Jones's (1980) out-group polarization findings and Marques and colleagues' (Marques et al., 1992; Marques & Yzerbyt, 1988; Marques et al., 1988) in-group polarization findings may be explained, in part, by expectancy-violation theory. From our view, whether researchers observe in-group or out-group evaluative extremity depends not on perceivers' group memberships, but on whether the target violates expectations for their respective categories. The findings of these studies, as well as those of other recent studies (Bettencourt et al., 1997; Biernat et al., 1996), suggest that evaluative polarization per se may not be the outcome of interest. Instead, from our perspective, extremity in affect-related evaluations is often a result of the surprise and the valence associated with a category-based expectancy violation, regardless of group membership.

Moreover, from our interpretation, the complexity-extremity and black-sheep hypotheses may underscore different parts of the evaluative process. For example, Linville and Jones's (1980) hypothesis emphasizes cognitive influences on evaluation, whereas Marques and colleagues' (Marques et al., 1992; Marques & Yzerbyt, 1988; Marques et al., 1988) hypothesis emphasizes motivational influences on evaluation. These viewpoints focus on different aspects of what may be a general process of evaluating targets with various group memberships. That perceivers have a more complex schema for their in-group than their out-group, for example, does not contradict the prediction that persons are often motivated to favor members of their in-group. In addition, in some situations, motivational factors may play the greater role in evaluations, whereas in other situations, cognitive factors may dominate. Together with expectancy-violation theory, these perspectives may help to inform us about the ways in which different motivational and cognitive factors affect evaluative extremity.

Several limitations of this work should be noted. First, because the studies were designed to test expectancy-violation theory, we purposely created targets who violated category-based expectancies. As such, our studies were not meant as a critical test of the three perspectives (expectancy violation, complexity-extremity, and black-sheep) and our findings are somewhat limited in their application to the complexity-extremity and black-sheep perspectives.

Second, there were a relatively low number of minority group members available for this study. Although several different strategies were used (e.g., attempting to recruit participants outside the Introductory Psychology pool, recruiting participants over several academic semesters), it was impossible given the racial composition of our Midwestern campus to obtain equal numbers of White and Black participants. Even so, our results show a similar, albeit somewhat weaker pattern of expectancy violation effects for Black participants as compared to Whites, and this pattern replicates previous work using only Black participants (Coleman et al., 1995).

Third, although generally supportive of expectancy-violation theory, the effects in Experiment 1 were small in some instances, a finding that may have several explanations. These effects may be attributable to the targets themselves. The ACT scores shown for the weak target applicant (75%) may have been too high, such that our weak targets were also somewhat positive. It should be noted, however, that Jackson et al. (1993) used similarly qualified weak targets to obtain expectancy-violation effects. Nevertheless, our findings suggest that lower ACT scores may elicit even stronger affective reactions to the weak targets. Another explanation for the relatively weak effects may involve the way in which affect-related evaluations and attributions were measured. According to Biernat and Manis (1994), standards of judgment may shift when the group of reference changes. In other words, it may have been that when participants were answering questions for a Black target, they used the stereotype for Blacks to "anchor" both ends of the numbered rating scale used to measure their responses. Thus, the endpoints of the scale may have represented slightly different meanings as participants evaluated different target groups, thus contributing to smaller differences between the targets overall. This perspective suggests that these affective reactions may be even stronger in more realistic settings when measurement on a numbered rating scale is not required.

Finally, that only one target per ethnic group category was used may reduce the generalizability of our findings. It should be noted, however, that expectancy-violation effects have been found in a wide range of studies using a wide range of target materials (Bettencourt et al., 1997; Coleman et al., 1995; Jackson et al., 1993; Jussim et al., 1987; Jussim et al.,1996).

In conclusion, evaluations of group members are likely to be affected by a confluence of factors (Jussim et al., 1987). As have other studies (e.g., Bettencourt et al., 1997; Biernat et al., 1996; Coleman et al., 1995; Jussim et al., 1987; Jussim et al., 1996), these studies show strong effects of the valence of the individuating information (academic credentials of the target) on affect-related evaluations. Even in the presence of this individuating information, however, the effects of category-based expectancy violations were revealed. These results suggest that evaluations may be more influenced by individuating than categorical information (Jussim et al., 1987; Jussim et al., 1996; Locksley et al., 1980; Locksley et al., 1982), except when individuating information is in direct

contrast with categorical expectancies. In the latter case, the more extreme evaluations of category-violating targets suggest that perceivers take into account categorical information when evaluating them. As such these results support recent theorizing that categorical and individuating processes simultaneously affect evaluative and judgmental processes (Kunda & Thagard, 1996; Nelson, Acker, & Manis, 1996).

ACKNOWLEDGMENTS

This research was supported by National Science Foundation Grant SBR9409921, and by a grant from the University of Missouri, both to B. Ann Bettencourt.

Preliminary reports of this research were presented at the annual meeting of the Society for the Psychological Study of Social Issues, August 1997 and June 1998.

We thank Sin-Myoung Lee, Rachel Phillips, and Scott Greathouse for their assistance in data collection.

REFERENCES

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62,* 893–912.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bassili, J. N., & Racine, J. P. (1990). On the process relationship between person and situation judgments in attribution. *Journal of Personality and Social Psychology, 59,* 881–890.

Bettencourt, B. A. (1997). *Effects of category-based expectancy violation on evaluations and motivated reasoning.* A grant application submitted to the National Science Foundation.

Bettencourt, B. A., Dill, K. E., Greathouse, S., Charlton, K., & Mulholland, A. (1997). Predicting evaluations of ingroup and outgroup members: The role of category-based expectancy violation. *Journal of Experimental Social Psychology, 33,* 244–275.

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology, 66,* 5–20.

Biernat, M., Vescio, T. K., & Billings, L. S. (1996). *Accounting for patterns of ingroup and outgroup judgment: Models of complexity, ambivalence, black sheep, and expectancy violation.* Unpublished manuscript.

Branscombe, N. R., Wann, D. L., Noel, J. G., & Coleman, J. (1993). In-group or out-group extremity: Importance of the threatened social identity. *Personality and Social Psychology Bulletin, 19,* 381–388.

Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology, 12,* 30–48.

Cedrus Corporation. (1991). *SuperLab* (Version 1.4) [Computer software]. Phoenix, AZ: Author.

Chaiken, S. D., & Bargh, T. A. (1993). Occurrence versus moderation of the automatic attitude activation effort: Reply to Fazio. *Journal of Personality and Social Psychology, 64,* 759–765.

Clary, E. G., & Tesser, A. (1983). Reactions to unexpected events: The naive scientist and interpretive activity. *Personality and Social Psychology Bulletin, 9,* 609–620.

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition: Vol. 1. Basic processes* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Coleman, L. M., Jussim, L., & Kelly, S. H. (1995). A study of stereotyping: Testing three models with a sample of Blacks. *Journal of Black Psychology, 21,* 332–356.

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology: Review of personality and social psychology* (Vol. 11, pp. 74–97). Newbury Park, CA: Sage.

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe–Crowne social desirability scale. *Educational and Psychological Measurement, 53,* 417–424.

Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist, 48,* 621–628.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Forgas, J. P. (1995). Strange couples: Mood effects on judgments and memory about prototypical and atypical relationships. *Personality and Social Psychology Bulletin, 21,* 747–765.

Graham, S., & Brown, J. D. (1988). Attributional mediators of evaluation, expectancy, and affect: A response time analysis. *Journal of Personality and Social Psychology, 55,* 873–881.

Greenberg, J., & Pyszczynski, T. A. (1982). Role of disconfirmed expectancies in the instigation of attributional processing. *Journal of Personality and Social Psychology, 40,* 31–38.

Hamilton, D. L., & Zanna, T. (1972). Differential weighting of favorable and unfavorable attributes in impression of personality. *Journal of Experimental Research in Personality, 6,* 204–212.

House, W. C., & Perney, V. (1974). Valence of expected and unexpected outcomes as a function of locus of goal and type of expectancy. *Journal of Personality and Social Psychology, 29,* 454–463.

Jackson, L. A., & Cash, T. F. (1985). Components of gender stereotypes and their implications for stereotypic and nonstereotypic evaluations. *Personality and Social Psychology Bulletin, 11,* 326–344.

Jackson, L. A., Sullivan, L. A., & Hodge, C. N. (1993). Stereotype effects on attributions, predictions, and evaluations: No two social judgments are quite alike. *Journal of Personality and Social Psychology, 65,* 69–84.

Jones, E. E. (1990). Expectancy effects. In R. C. Atkinson, G. Lindzey, & R. F. Thompson (Eds.), *Interpersonal perception* (pp. 77–111). New York: Freeman.

Jussim, L., Coleman, L. M., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality and Social Psychology, 52,* 536–546.

Jussim, L., Fleming, C. J., Coleman, L. M., & Kohberger, C. (1996). The nature of stereotypes II: A multiple-process model of evaluations. *Journal of Applied Social Psychology, 26,* 283–312.

Kernahan, C., & Bettencourt, B. A. (1998). *Motivated processing of Black and White targets: The situation of choice versus assignment.* Manuscript submitted for publication.

Kernahan, C., Greathouse, S. A., & Bettencourt, B. A. (1996, May). *Extreme evaluations as a function of expectancy violation and humanitarian–egalitarian values.* Poster session presented at the sixty-eighth annual meeting of the Midwestern Psychological Association, Chicago.

Klein, S. B., Loftus, J., & Burton, H. A. (1989). Two self-reference effects: The importance of distinguishing between self-descriptiveness judgments and autobiographical retrieval in self-referent encoding. *Journal of Personality and Social Psychology, 56,* 853–865.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel constraint-satisfaction theory. *Psychological Review, 103,* 284–308.

Lambert, A. J. (1995). Stereotypes and social judgment: The consequences of group variability. *Journal of Personality and Social Psychology, 68,* 388–403.

Linville, P. W., & Jones, E. E. (1980). Polarized appraisals of outgroup members. *Journal of Personality and Social Psychology, 38,* 689–703.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology, 39,* 821–831.

Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology, 18,* 23–42.

MacDowell, K. A., & Mandler, G. (1989). Constructions of emotion: Discrepancy arousal and mood. *Motivation and Emotion, 13,* 105–124.

Macrae, C. N., & Shepherd, J. W. (1991). Categorical effects on attributional inferences: A response-time analysis. *British Journal of Social Psychology, 30,* 235–245.

Marques, J. M, Robalo, E. M., & Rocha, S. A. (1992). Ingroup bias and the "black-sheep" effect: Assessing the impact of social identification and perceived variability on group judgments. *European Journal of Social Psychology, 22,* 331–352.

Marques, J. M., & Yzerbyt, V. Y. (1988). The black-sheep effect: Judgmental extremity towards ingroup members in inter- and intra-group situations. *European Journal of Social Psychology, 18,* 287–292.

Marques, J. M., Yzerbyt, V. Y., & Lyens, J. (1988). The black-sheep effect: Extremity of judgments towards ingroup members as a functions of group identification. *European Journal of Social Psychology, 18,* 1–16.

McIntosh, P. (1992). White privilege and male privilege: A personal account of coming to see correspondences through work in women's studies. In M. L. Anderson & P. H. Collins (Eds.), *Race, class and gender: An anthology* (pp. 70–81). Belmont, CA: Wadsworth.

Nelson, T. E., Acker, M., & Manis, M. (1996). Irrepressible stereotypes. *Journal of Experimental Social Psychology, 32,* 13–38.

Newman, L. S. (1993). How individuals interpret behavior: Idiocentrism and spontaneous trait inference. *Social Cognition, 11,* 243–269.

Roediger, D. R. (1998). *Black on White: Black writers on what it means to be White.* New York: Schocken.

Smith, E. R., & Miller, F. D. (1979). Attributional information processing: A response time model of causal subtraction. *Journal of Personality and Social Psychology, 37,* 1723–1731.

Smith, E. R., & Miller, F. D. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general model. *Journal of Personality and Social Psychology, 44,* 492–505.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics.* New York: Harper & Row.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93,* 239–257.

Tukey, J. W. (1990). Data-based graphics: Visual display in the decades to come. *Statistical Science, 5,* 327–339.

Verplanken, B., Hofstee, G., & Janssen, H. J. W. (1998). Accessibility of affective versus cognitive components of attitudes. *European Journal of Social Psychology, 28,* 23–35.

Wilcox, R. R. (1995). *Statistics for the social sciences.* San Diego, CA: Academic.

Copyright of Basic & Applied Social Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.