An On-Line Look at Automatic Contrast and Correction of Behavior Categorizations and Dispositional Inferences

Stephanie J. Tobin
Gifford Weary
Ohio State University

The current study examined on-line behavior recategorization as a mechanism underlying corrections for contextual influences in dispositional inferences. After watching an initial comparison video that portrayed either a successful or unsuccessful performance on a spatial ability task, cognitive load and no load participants watched and made real-time ratings of a target performance. The comparison video was expected to exert a contrastive influence on participants' automatic impressions of the performance (behavior categorizations) and the child's intelligence (dispositional inferences). Load participants' on-line and post-video performance and ability ratings showed this expected effect, as did no load participants' initial on-line performance ratings. However, no load participants' later on-line and post-video ratings did not. These findings support the notion that corrections for contextual influence can occur at the level of behavior identification as perceivers encode behavioral cues.

Keywords: *dispositions; on-line inferences; behavior recategorization; comparison standards*

Assessing the underlying dispositions of others is of primary importance if one wishes to explain, predict, and control the social world (Heider, 1958). Thus, it is no surprise that researchers have devoted a great deal of attention over the years to understanding how people draw dispositional inferences, where they might go wrong, and how they might correct for, or altogether avoid, potential mistakes (e.g., Gilbert, 1989; Heider, 1958; Jones & Davis, 1965; Kelley, 1972; Trope, 1986). Our research is in this tradition. More specifically, we focus here on the implicit effects of contextual information on dispositional judgments and on a mechanism of correction heretofore neglected by attribution researchers: behavior recategorization. Before describing our research, however, we first review briefly the major contemporary models of the dispositional inference process.

DISPOSITIONAL INFERENCE MODELS

The contemporary models divide the process of making inferences about the causes of others' behaviors into two stages: identification and dispositional inference (Gilbert, Pelham, & Krull, 1988; Trope, 1986, 1998). In the identification stage, perceivers categorize a person's behavior in trait-relevant terms (e.g., he performed intelligently). They then use this categorization in the dispositional inference stage and often entertain the initial hypothesis that a correspondent disposition (e.g., he is intelligent) was responsible for the observed behavior. If motivated and able, they subsequently correct this dispositional inference for any plausible, alternative causal factors (e.g., task ease).

The stage models also suggest that two basic kinds of errors can occur in the dispositional inference process: identification and inference errors (Trope, 1986, 1998). Identification errors result when situational information exerts an undue influence on perceivers' categorizations of another's behavior. This is particularly likely to hap-

Authors' Note: This research was supported in part by National Science Foundation Grant SBR-9631858 awarded to Gifford Weary. We thank Jill Jacobson for suggesting the on-line rating procedure used in this study and Blair Jarvis for designing the necessary software. We also thank Jill Jacobson, Darcy Reich, Aaron Wichman, and two anonymous reviewers for their helpful comments on an earlier version of this article. Correspondence concerning this article should be addressed to Stephanie Tobin or Gifford Weary, Department of Psychology, Ohio State University, 142 Townshend Hall, 1885 Neil Avenue, Columbus, OH 43210; e-mail: tobin.31@osu.edu or weary.1@osu.edu.

PSPB, Vol. 29 No. 10, October 2003 1328-1338
DOI: 10.1177/0146167203254611

pen when behavioral cues are ambiguous, or open to different categorizations, and when contextual information precedes observation of the behavior (Trope & Alfieri, 1997; Trope, Cohen, & Alfieri, 1991; Trope, Cohen, & Maoz, 1988). The behavior of a weeping woman, for example, likely would be construed differently depending on whether it were observed within the context of a funeral or a wedding (i.e., she may be seen as shedding tears of sorrow at a funeral but tears of joy at a wedding). Certainly, such knowledge of the situation oftentimes can be essential in achieving an accurate understanding of another's behavior. However, it also can bias behavior interpretations and initial dispositional inferences. It can cause perceivers to see a behavior and corresponding disposition as more extreme than they actually are.

Inferential errors occur when perceivers fail to consider carefully all plausible alternative causes (e.g., the situation) of an observed behavior. Returning to our earlier example, if perceivers were to fail to take the funeral setting into account, they might well have concluded that the woman was depressed, when in fact she was not (Gilbert & Malone, 1995; Trope, 1986). Such a thorough analysis of alternative explanations requires a fair amount of motivation and the devotion of cognitive resources. It is, therefore, not surprising that inferential errors are more likely when perceivers are under cognitive load or when their motivation is low (Gilbert et al., 1988; Gilbert, Krull, & Pelham, 1988; Trope & Alfieri, 1997).

Is one type of error more difficult to address than the other? The available evidence (Gilbert & Malone, 1995; Trope, 1986; Trope & Alfieri, 1997) suggests that inferential errors are easier to correct than are errors of identification. The latter occur automatically and appear to the perceiver to reflect the objective properties of the stimulus or person (Trope, 1986, 1998; Trope & Alfieri, 1997; Trope & Gaunt, 1999). For this reason, perceivers are more likely to adjust their dispositional inferences for plausible situational explanations of behavior (i.e., the woman appeared depressed because she was at a funeral) than they are to remove situational influences from their behavior categorizations (i.e., the woman's behavior appeared more depressed than it really was because it occurred at a depressing event).

Trope and Alfieri (1997) recently offered support for the idea that correction of behavior categorizations does not occur, even under optimal processing conditions. In their study, they exposed participants, who either were under load or no load, to situational information indicating that an evaluator had been pressured to provide either a positive or negative assessment of a job candidate. Next, participants were given the evaluator's ambiguous assessment of the job candidate. Trope and Alfieri found that although no load participants were able to correct their judgments of the evaluator's true opinion of the candidate (dispositional inference) for the situational demands, they were unable to adjust their categorizations of the evaluation itself; that is, they still assimilated the ambiguous assessment to the situational demands.

THE CURRENT RESEARCH

Although the findings of the above-mentioned study provide some support for the notion that recategorization is not possible, other researchers have argued that under certain circumstances, perceivers may be able to recategorize observed behavior (Kunda, 1998; Weary, Reich, & Tobin, 2001; Weary, Tobin, & Reich, 2001). Indeed, much of the evidence for this argument has come from our lab.

In our studies, participants typically are asked to view a videotape of a child performing several spatial ability tasks and then to rate the success of his performance (behavior categorization) and his level of intelligence (dispositional inference). The results generally show that contextual information about the target child's previous performance or about another child's performance influences load participants' behavior categorizations and dispositional inferences in either an assimilative or contrastive manner, with the direction of the effect depending on how the information is categorized.[1] The results also show, however, that the judgments of no load participants are adjusted for the contextual information; that is, their behavior categorizations and dispositional inferences show either no or reverse effects (e.g., in the direction of corrective assimilation when load participants show contrast) of the contextual performance.

Although our prior research suggests that behavior recategorization may be possible under certain circumstances, it is important to note that our studies assessed behavior categorizations after an entire behavioral episode had been encoded (Weary, Reich, & Tobin, 2001; Weary, Tobin, & Reich, 2001). Such a method leaves open the possibility that perceivers first corrected their dispositional inferences in light of the contextual bias and then later merely adjusted their behavior categorizations to be consistent. More important, such a method also leaves open the possibility that the no load perceivers managed to avoid the influence of the context altogether, that there was no initial bias for which corrections needed to be undertaken. A method that is able to demonstrate early influence and later correction within the same group of participants would provide stronger evidence for a correction stage.

With this goal in mind, the current study sought to examine perceivers' continuous behavior categorizations as they encoded the target's behaviors. More specifically, we assessed on-line behavior categorizations by having participants use a joystick to rate continuously a videotaped performance of a child completing several spatial ability tasks. We manipulated contextual information by first showing participants a video that depicted a very successful or unsuccessful performance of a different child performing similar tasks. Furthermore, we placed half of our participants under cognitive load while they watched and rated the main video.

Based on past research, we expected that the prior performance would be seen as distinct from the target's current performance and would, as a result, serve as a comparison standard (Martin & Seta, 1983; Weary, Tobin, & Reich, 2001). Therefore, we expected to see contrast from it in participants' on-line categorizations of the main target's behavior under cognitive load conditions. When participants were under no load, however, we expected the contextual influence of the prior performance on their on-line behavior categorizations to change over time; that is, we expected to see contrast from contextual information initially, followed by a recategorization of behavior. This recategorization was expected to result from an awareness of bias arising from the phenomenological experience of surprise (Kunda, 1998; Wilson & Brekke, 1994) that often accompanies unexpected events (e.g., the child did much better than I thought he would). In addition to obtaining participants' on-line ratings, we also obtained final assessments of their postvideo behavior categorizations and dispositional judgments. We expected to replicate our past findings on these reported target inferences. Specifically, we expected to see contrast from contextual information under load and no contrast under no load conditions.

METHODS

Video Piloting

For the current study, we required comparison performances that were clearly better and clearly worse than the target's performance in the main video. To this end, we created two new comparison videos: one showed a child successfully completing two block designs from the Stanford-Binet test in a total of 56 s, and the other showed the same child failing to complete a single design in 74 s.[2] The main video, by contrast, showed a different child completing three block designs in a total of 211 s (as used in Weary, Reich, & Tobin, 2001; Weary, Tobin, & Reich, 2001). All videos were silent and the performance outcomes were partially obscured. To assure that the comparison videos were categorized correctly, however, we provided participants with additional information about those performances. Specifically, before participants watched the successful and unsuccessful comparison videos, we informed them that the child had been either very successful or not successful in matching all designs within the time limits, respectively.

We examined participants' perceptions of the three videos in a pilot study. Thirty-eight men ($N = 11$) and women ($N = 27$) enrolled in an introductory psychology class were randomly assigned to view the unsuccessful comparison video, the successful comparison video, or the main video. Participants were first given the opportunity to practice using the joystick by moving it from left to right for 45 s. During this practice period, they were able to observe that their joystick movements controlled a pointer that moved along a 9-point scale on the computer screen (1 = *all the way left*, 9 = *all the way right*). Following this practice period, participants were told that they would watch a video of a child performing a spatial ability task and that their goal was to figure out how intelligent the child was. They were instructed to use the joystick to rate continuously the overall quality of the performance in the video on a scale of 1 (*very unsuccessful*) to 9 (*very successful*). After watching and rating one of the videos, participants completed several additional items. Specifically, they rated on 9-point scales the difficulty of the spatial ability task (*not at all difficult* to *very difficult*), the child's likability (*not at all likable* to *very likable*), the ease with which they had been able to tell how well the child had performed (*not at all easy* to *very easy*), the clarity of the child's performance (*not at all unclear* to *very unclear*), the clarity of the task's difficulty (*not at all unclear* to *very unclear*), and the difficulty of ascertaining the child's performance outcome (*very easy* to *very difficult*).

A one-way analysis of variance (ANOVA) conducted on participants' average on-line ratings revealed that the videos differed in the level of success that they portrayed, $F(2, 34) = 9.84$, $p < .001$. Planned t test comparisons revealed that compared to the main video ($M = 5.46$), the performance in the successful comparison video was viewed as significantly better ($M = 6.01$), $t(34) = -2.03$, $p < .05$, and the performance in the unsuccessful comparison video was viewed as significantly worse ($M = 4.71$), $t(34) = 2.68$, $p < .05$.[3] Analyses of the other postvideo ratings revealed no significant video differences (ps > .17).

Participants

Participants in the main study were 97 men ($N = 45$) and women ($N = 52$) enrolled in an introductory psychology class. The data for 1 participant were excluded because she reported having difficulty controlling the joystick. Up to 7 participants were run in a given session, but each participant was assigned to an individual cubicle. All participants were randomly assigned to conditions in this 2 (comparison standard: unsuccessful, successful) × 2 (cognitive load: no load, load) between-subjects factorial design.

Procedure

A female experimenter greeted the participants and explained that the study would be conducted on computers. Participants then were led to individual cubicles and were told to begin the experiment, which was run by *MediaLab* (Jarvis, 2000).

After answering some demographic questions, participants read that they would be watching a videotape of a child performing a spatial ability task and that afterward they would be asked to give their impressions of the performance. They also learned that they would use a joystick to make their ratings of the child. Next, participants were given two practice tasks to become familiar with the joystick rating procedure. First, they practiced moving the joystick from left to right for 20 s as the joystick position appeared on the computer screen on a scale of 1 (*all the way left*) to 9 (*all the way right*). Then, they used the joystick to rate a dynamic target; specifically, they used it to rate on a scale of 1 (*very small*) to 9 (*very big*) the size of a square that grew and shrank in size over the course of a minute.

After the two practice periods, participants received additional instructions about the upcoming videotape task. They were told that they could select one of four videotapes, each of which showed a different child completing different spatial ability tasks. The tasks were described as puzzles that were part of a general intelligence test. Participants were told that some children performed their tasks quite successfully, whereas others were less successful. They then were given the following instructions:

> It will be your task to figure out HOW INTELLIGENT THE CHILD IS in general. As you know, very intelligent people sometimes appear less intelligent because they are performing a very difficult task and less intelligent people sometimes appear very intelligent because they are performing a very simple task. We do not want you to tell us how intelligent the child merely appears. Instead, we want you to watch the child on the tape and figure out how intelligent you think the child is in general.

Participants next were informed that they would be asked to use the joystick throughout the video to indicate the overall quality of the child's performance on the various puzzles and that their continuous ratings of the performance would appear on a scale below the video. These ratings ranged from *very unsuccessful* to *very successful*. Participants next chose one of the four ostensibly different videotapes to watch as the main video. In reality, all participants watched the same main video but we wanted them to think that that it was selected from a range of performances. After selecting a tape, they were given some additional information about their selection. First, they were told,

> We find that people feel more comfortable answering questions about the performance in the main video when they are familiar with what is typically involved in a spatial ability task. Since most college students do not have much experience with spatial ability tasks, we usually show participants a short example video clip to familiarize them with the task before they watch the main performance.

Participants then read some additional comments about the child in the main video. Specifically, they were told that the child in the main video was 9 years old and that he was part of a family testing program and happened to be the brother of the child in the example video that they were about to watch.

Comparison standard video. Prior to viewing the example, or comparison video, participants received the following instructions with either the word *VERY* or *NOT* inserted before successful:

> When you click continue, a video clip of a child performing some spatial ability puzzles will play. These puzzles are part of a general intelligence test and the child's task is to put together the blocks so that they match a design provided by the examiner. You will not be able to see the design that the child is trying to match because of the camera angle in this particular video. However, this child was VERY (NOT) successful in matching all designs within the time limits.

Half of the participants further were instructed to use the joystick to rate the overall quality of the performance in the comparison video.[4] All participants then watched (and half also rated) either the successful or unsuccessful comparison tape.

Cognitive load. After watching one of the comparison standard videos, all participants were reminded that while watching the main video, they were to use the joystick to rate the overall quality of the performance in the video and that their task was to figure out how intelligent the child was. Participants in the load condition also were told that an 8-digit code number would appear on the next screen right before the videotape started. They were to rehearse this number while watching the video, without writing it down anywhere. The 8-digit number then appeared for 20 s and was followed by the main video.

Dependent measures. Joystick ratings of the success of the performance in the main videotape served as the primary dependent measure in the current study. As participants watched and rated the main video, the position of the pointer on the scale was recorded every 100 ms. These ratings then were averaged for each second and written to a data file. We also were interested in participants' postvideo assessments of the target's behavior and level of ability. Accordingly, after watching the main video, participants completed items that assessed their behavior categorizations ("The success of the children's performance on the four videotapes you chose from differed. What level of performance was depicted in the tape you watched?" 1 = *very unsuccessful* to 9 = *very successful*), dispositional inferences ("How would you rate the current spatial ability of the child in the video?" 1 = *very low* to 9 = *very high*), and ability-irrelevant impressions of the child ("How likable is the child in the video?" 1 = *not at all likable* to 9 = *very likable*; and "How would you rate this child's social skills?" 1 = *very poor* to 9 = *very good*). Following these items, participants in the load condition recorded the digits they had been rehearsing.

Next, all participants completed a manipulation check item that asked them to rate on a scale of 1 (*very unsuccessful*) to 9 (*very successful*) how successful the performance had been in the short video (i.e., the comparison video) that they had watched before the main one. They also rated whether watching the first video made their initial impressions of the child in the main video more positive or more negative than they would have been without having watched the first video (–4 = more negative, +4 = more positive).

Following these items, participants completed six surprise multiple-choice questions that tested their attention to detail in the main video. Specifically, participants indicated the number of test items in the video, the color of the child's shirt, what was in the background, the color of the test administrator's hair, the color of the blocks, and the relation of the child in the video to the child in the short video they watched earlier. Finally, all participants were fully debriefed and thanked.

RESULTS

Recall

Digit recall. Researchers recommend examining the overall recall rate on a cognitive load task to ensure that the task is difficult enough to present a cognitive demand for participants (Bargh & Chartrand, 2000; Gilbert & Hixon, 1991). The overall rate of digits correctly recalled and placed in the current study (M = 80%) was comparable to that found in other studies (Weary, Reich, & Tobin, 2001; Weary, Tobin, & Reich, 2001) and indicated that the task was not too easy for participants.

In addition, researchers generally suggest that participants with extremely low recall be excluded. Such participants likely are not engaged in the secondary memory task and, thus, are not under load (Bargh & Chartrand, 2000; Gilbert & Hixon, 1991; Weary, Reich, & Tobin, 2001; Weary, Tobin, & Reich, 2001). Accordingly, we excluded the nine participants who correctly recalled fewer than half of the digits. For the remaining participants, there was no effect of comparison standard condition on the number of digits correctly recalled and placed ($p = .69$).

Video recall. The average number of video items correctly recalled in the current study was 4.71 out of 6. This rate is very similar to that found in previous studies using the same video (Weary, Reich, & Tobin, 2001; Weary, Tobin, & Reich, 2001) and suggests that introducing the on-line rating procedure did not interfere with encoding the details of the video. Consistent with our exclusion rule for digits recalled, we excluded the one participant who correctly recalled fewer than half of the details about the videotape. For the remaining participants, there were no effects of comparison standard or load conditions on video recall (ps > .37)

Comparison Videos

A 2 (comparison video) × 2 (cognitive load) ANOVA on participants' postvideo ratings of how successful the comparison video was revealed only the expected main effect of comparison video, $F(1, 81) = 30.40$, $p < .001$. Participants rated the unsuccessful video as less successful (Ms = 3.22 vs. 6.15).

Main Video

Participants provided continuous success ratings of the child's performance on three block design items in the main video so that we could examine over time the possible contrastive effects of the comparison videos and, among no load participants, potential on-line behavior recategorization. As one examines these data, displayed in Figure 1, it is clear that comparison video condition had an impact on ratings of the main performance. Participants who had seen the unsuccessful versus successful comparison video tended to rate the main performance as more successful. However, the degree of contrast differs across video clips (i.e., test items) as a function of load. The lack of contrast in Video Clip 3 among no load participants indicates that these participants were able to engage in recategorization, removing the early influence of the comparison video from their ratings of the main performance.

In addition to these patterns, there also are within-video clip patterns exhibited by load and no load participants alike. These patterns mirror the child's performance progress in each video clip: the child begins and then completes a puzzle. Because success is difficult to determine in the beginning of each clip (i.e., when the

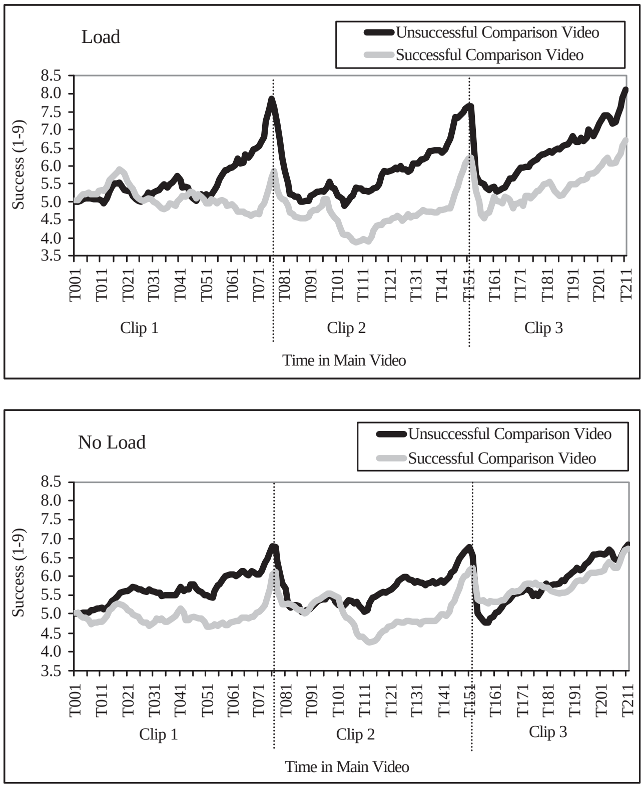

Figure 1 On-line ratings of the main video as a function of comparison video and cognitive load.
NOTE: Each video clip showed the child beginning and completing a puzzle.

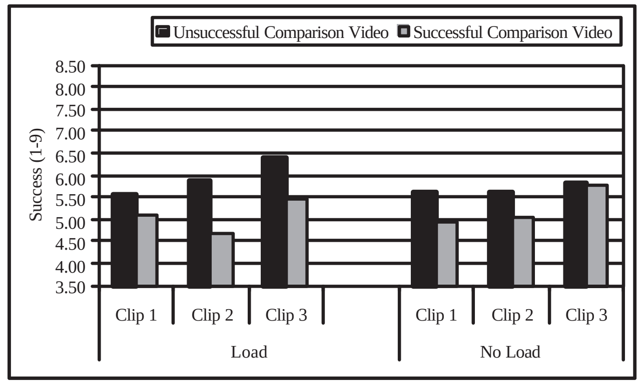

Figure 2 Average on-line ratings of the main video as a function of comparison video, load, and video clip.

child has placed only a few blocks in position), participants' performance ratings hover around the scale midpoint. At the end of each clip, the completion of a design signifies at least some level of success, and participants' performance ratings for the child are higher as a result.

To examine whether the degree of contrast differed significantly across the test items as a function of load, we averaged on-line ratings within video clip and then conducted a 2 (comparison video) × 2 (load) × 3 (clip) mixed-model ANOVA with repeated measures on the last factor. In addition to significant main effects of video clip, $F(2, 162) = 19.06$, $p < .001$, and comparison video, $F(1, 81) = 13.55$, $p < .001$, the analysis revealed a significant Comparison Video × Load × Video Clip interaction effect, $F(2, 162) = 4.15$, $p < .05$. Planned comparisons revealed the predicted effects: Load participants showed significant contrast from the comparison video during all three video clips, ts(162) = 2.03, 5.46, and 4.41 for clips 1, 2, and 3, respectively, ps < .05, but no load participants only showed significant contrast during the first two video clips, ts(162) = 3.52 and 3.02 for clips 1 and 2, respectively, ps < .05; $t(162) = 0.25$ for clip 3, $p > .80$ (see Figure 2). No load participants' on-line ratings on video clip 3 were indicative of behavior recategorization.[5]

Final Target Judgments

We next examined whether the on-line rating procedure produced any differences in participants' postvideo ratings of the target's performance and ability levels. Recall that in past research employing the same procedures used herein, we have obtained inferential contrast under load and correction for contrast effects under no load conditions.

First, we examined participants' categorizations of the main target's behavior. A 2 (comparison video) × 2 (cognitive load) ANOVA revealed a significant main effect of comparison video, $F(1, 81) = 5.84$, $p < .05$, on performance ratings such that those participants in the unsuccessful ($M = 7.34$) compared to successful ($M = 6.42$) comparison video conditions rated the main performance as significantly more successful. The interpretation of this main effect, however, was qualified by the presence of a significant Comparison Video × Load interaction effect, $F(1, 81) = 5.87$, $p < .05$. Planned comparisons revealed a significant contrast effect under load: Compared to those participants who had viewed the successful comparison video ($M = 6.20$), those exposed to the unsuccessful one ($M = 8.05$) perceived the main performance as significantly better, $t(81) = 3.34$, $p < .05$. Under no load conditions, however, participants' behavior perceptions of the main target were unaffected by the comparison performances (Ms = 6.64 vs. 6.63 for the unsuccessful and successful comparison video conditions, respectively), $t(81) = 0.01$, $p > .99$.

We next examined participants' postvideo dispositional inferences of the child's ability level. A 2 (comparison video) × 2 (cognitive load) ANOVA revealed a significant main effect of comparison video on ability ratings such that those who viewed the unsuccessful ($M = 7.09$) compared to successful ($M = 6.04$) comparison video rated the child as possessing significantly higher ability, $F(1, 81) = 7.72$, $p < .05$. The interpretation of this main effect, however, was qualified by the presence of a significant

Comparison Video × Load interaction effect, $F(1, 81) = 4.05$, $p < .05$. Planned comparisons revealed a significant contrast effect under load: Participants who viewed the unsuccessful ($M = 7.62$) compared to successful ($M = 5.80$) comparison video viewed the main child as significantly higher in ability, $t(81) = 3.30$, $p < .01$. Under no load conditions, however, participants' perceptions of the main child's ability were not significantly affected by the comparison video (Ms = 6.57 vs. 6.27 for the unsuccessful and successful comparison video conditions, respectively), $t(81) = 0.58$, $p > .56$.

Mediational Analyses

Although analyses of participants' later on-line and postvideo target ratings revealed the same pattern of results (contrast among load participants, correction among no load participants), a more precise statement of the process by which participants draw their dispositional inferences requires the support of mediational analyses. Theoretically, dispositional inferences should be based on categorizations of the observed behavior, because behavior is the raw data on which higher order social inferences are based (Heider, 1958). Among our load participants, this relationship should have been fairly straightforward, because their categorizations were influenced by the comparison video during all three video clips; their postvideo behavior categorizations and dispositional inferences should have been consistent with their early (and, of course, later) on-line impressions. This consistency of early on-line impressions and postvideo behavior categorizations, however, was not expected for no load participants; their behavior categorizations were expected to be influenced early on by, but to be corrected later for, the contextual influence of the comparison videos. Their final, postvideo behavior categorizations and dispositional inferences, then, should have been based on their corrected on-line impressions.

To examine what participants based their dispositional inferences on, we first created two on-line indices: one was the average of participants' on-line ratings prior to the point at which no load participants showed recategorization (Video Clips 1 and 2) and the other was the average of participants' on-line ratings after this point (Video Clip 3). We next examined whether these on-line indices predicted participants' final, postvideo behavior categorizations and, in turn, their dispositional inferences. We conducted these analyses within load condition because the process was expected to differ as a function of cognitive resources (i.e., only no load participants could reasonably be expected to correct their categorizations).

Starting with the load condition, and following the steps outlined by Baron and Kenny (1986), we first regressed dispositional inferences on effects-coded gender (–1 = male, +1 = female), initial on-line ratings (Video Clips 1 and 2), and later on-line ratings (Video Clip 3). Gender ($\beta = .58$, $t = 2.38$, $p < .05$) and initial on-line ratings ($\beta = 1.18$, $t = 3.42$, $p < .01$) significantly predicted dispositional inferences; not surprisingly, later on-line ratings did not significantly predict dispositional inferences when the variance attributable to initial ratings was statistically controlled ($\beta = -.02$, $t = -0.06$, $p > .95$). Next, we regressed the proposed mediator (final, postvideo behavior categorizations) on effects-coded gender, initial on-line ratings, and later on-line ratings. Gender ($\beta = .50$, $t = 2.33$, $p < .05$) and initial on-line ratings ($\beta = 1.33$, $t = 4.39$, $p < .001$) significantly predicted final, postvideo behavior categorizations, whereas later on-line ratings did not ($\beta = -.10$, $t = -0.37$, $p > .71$), again indicating that participants' final behavior categorizations were adequately explained by their initial impressions of the performance. We next examined whether final behavior categorizations predicted dispositional inferences when controlling for gender and on-line impressions. This was the case ($\beta = .82$, $t = 6.18$, $p < .001$) and, in addition, gender and on-line impressions became nonsignificant (ps > .36), indicating that the effect of early on-line impressions on dispositional inferences was mediated by final behavior categorizations. A Sobel test revealed that the drop in predictive power of early on-line impressions was significant when final behavior categorizations were included in the model, $Z = 3.61$, $p < .001$. These findings support the notion that load participants based their dispositional inferences on their final behavior categorizations, which were established early on in the video.

Next, we conducted the same set of analyses with the no load condition participants. Both initial ($\beta = 1.03$, $t = 3.35$, $p < .01$) and later on-line ratings ($\beta = .53$, $t = 1.99$, $p = .05$) significantly predicted dispositional inferences. Only later on-line ratings significantly predicted postvideo categorizations ($\beta = .86$, $t = 2.87$, $p < .01$). In addition, final behavior categorizations predicted dispositional inferences when gender and on-line impressions were included in the model ($\beta = .60$, $t = 6.01$, $p < .001$). Later on-line impressions no longer predicted dispositional inferences when final categorizations were included in the model ($\beta = .01$, $t = 0.03$, $p > .97$), indicating that the effect of later on-line impressions on dispositional inferences was mediated by final behavior categorizations. A Sobel test revealed that this drop in predictive power of later on-line impressions was significant, $Z = 2.62$, $p < .01$. In addition, early on-line impressions still exerted an unexpected direct effect on dispositional inferences ($\beta = .75$, $t = 3.27$, $p < .01$) when final behavior categorizations were controlled, suggesting that, at some level, early impressions did exert some influence, at least on dispositional inferences. Aside from this one unexpected finding, these analyses support the notion that behavior categorizations are the basis for dispositional inferences, even when they undergo correction.[6]

Awareness of Bias

To be able to correct for any contextual influences on behavior categorizations, participants must be aware that their judgments have been contaminated, know the direction of bias, and have the requisite resources available to correct their judgments (Wegener & Petty, 1997). Were our participants aware of the influence that the comparison video had on their perceptions of the main performance? Participants rated on a –4 (*more negative*) to +4 (*more positive*) scale whether the first video made their initial impressions of the child in the main video more positive or negative than they would have been without having watched the first video. We converted ratings to a 1 to 9 scale and analyzed the effects of comparison standard and load with a two-way ANOVA. Results revealed only a main effect of comparison video such that compared to participants who watched the successful comparison video ($M = 4.77$), those who watched the unsuccessful one ($M = 5.65$) thought it had made their impressions somewhat more positive, $F(1, 81) = 3.76$, $p = .056$. Thus, when asked, load and no load participants were able to report the direction of influence of the comparison video on their early categorizations of the target's performance. The above-reported analyses of on-line and postvideo ratings, however, revealed that only no load participants were able to adjust their inferences accordingly during the performance.

Ancillary Measures

Last, we examined whether the effects of our manipulations were specific to behavior categorizations and dispositional inferences or whether they also appeared on more general, ability-irrelevant impressions of the child. If contrast under load and correction under no load are limited to our focal measures, they more likely would be due to the use of the comparison performance as a standard for evaluating the main performance than to a more general affective reaction to the child. As expected, our manipulations had no significant effects on participants' perceptions of how likable or socially skilled the child was (ps > .07).

DISCUSSION

Using an on-line rating procedure, the current study examined automatic contrast in behavior categorizations and on-line behavior recategorization as a mechanism of correction in the dispositional inference process. Findings revealed that although participants' initial behavior categorizations were contrasted from accessible comparison standards, only those with the available cognitive resources were able to remove this influence from their judgments as time went on. Specifically, although both load and no load participants' early perceptions of the target performance were contrasted away from the comparison performance they had viewed earlier, only no load participants were able to overcome this influence during the last portion of the main video and in their final target judgments.

Behavior Recategorization in the Dispositional Inference Process

The current study provides direct evidence in support of Weary and her colleagues' prior assertions that on-line behavior recategorization can occur in the dispositional inference process (Weary, Reich, & Tobin, 2001; Weary, Tobin, & Reich, 2001). Why have we been able to find evidence of behavior recategorization, or correction, when others have not? As we have noted in previous articles, the answer most probably lies in procedural differences between studies reported by various researchers interested in behavior categorization processes.

More specifically, studies that have failed to find evidence of behavior recategorization (i.e., Trope & Alfieri, 1997) have provided participants with specific and prestructured information about situational inducements present in the behavioral context. Such information might well have served to disambiguate target behaviors. In our studies, we have provided participants with no such information. Indeed, they have had to infer not only the specific level of the target's performance but also the presence and nature of any relevant situational factors. The information provided to our participants about the behavioral context, then, likely has left considerable room for potential categorizations (i.e., a more or less successful performance) of the observed performance. Leaving open alternative categorizations likely makes it easier for participants to recategorize target behaviors. Supporting this notion, Trope and Sikron (as cited in Trope & Gaunt, 1999) found that activating alternative interpretations of ambiguous behavior through a priming procedure wiped out the usual assimilative effect of situational inducements on behavior categorizations in their paradigm.

The degree to which the available information constrains behavior categorization may be one factor that determines whether recategorization takes place. Another factor concerns the nature of the behavioral cues. Trope and his colleagues typically have examined contextual influences on ambiguous behaviors, defined as behaviors that are "highly associated with different categories" (Trope, 1989, p. 134). Although our behavioral episode can be categorized as more or less successful, it is not highly associated with success and failure.

Rather, the target's performance could better be characterized as vague or "weakly associated with the various categories" (Trope, 1989, p. 134). Trope (1989) suggests that vague behaviors may be less influenced by contextual factors but may be more likely to give rise to an awareness of contextual influence because their link to any particular category is weaker. However, he also expresses doubt that such an awareness would necessarily stop perceivers from relying on the contextually influenced categorization in the inferential stage of processing. Our findings should help to mitigate such general doubt. Future research, though, should try to provide more direct evidence for the moderating role that type of contextual information and type of behavior may play in the recategorization process.

One final factor that clearly influenced recategorization in the current study was the amount of behavioral information. We found that no load participants were influenced by accessible contextual information for a substantial portion of the video before engaging in recategorization. This highlights the importance of the length of stimulus exposure in behavior recategorization. Although having to make a final summative judgment after viewing only the first minute of the video could have sparked correction, it also is possible that correction would only have been evident after viewing the full 3 min of the video. It is likely that in addition to possessing sufficient levels of motivation and ability, perceivers also must be exposed to a fair amount of incoming behavioral information for recategorization processes to occur.

What Drives Recategorization?

We have suggested that in our paradigm, adjustments to behavior categorizations among no load participants resulted from their awareness of the impact of a perceived contextual bias. Indeed, we found some evidence that participants were aware of the influence of the first video on their initial impressions of the child in the main video: Those who initially viewed an unsuccessful versus successful performance thought it had made their initial impressions of the child more positive. But how exactly did participants go about adjusting on-line behavior categorizations? One possibility is that during the last task in the video, participants made direct adjustments to their ratings of the behavior in light of their awareness of the discrepancy between expected and actual performance; that is, because the behavior "seemed" different than expected, participants might have adjusted their categorizations for the contrastive influence of their expectancies.

A second possibility, however, is that they first adjusted not their behavior perceptions but their expectancies (i.e., comparison standards) in light of the target's performance on the first two tasks. Such an updating of expectancies could be the result of either a passive process of adaptation (Helson, 1964), or it could be based on an awareness that one's expectancies, or comparison standards, are unreasonable (Gilbert, Giesler, & Morris, 1995). Either way, such an adjustment would effectively reduce the discrepancy between expectancies and incoming behavioral cues if the actual behavior (i.e., level of performance) remained consistent over time. The lessened impact of the comparison videos on expectancies and the reduced discrepancy between expectancy and behavior should result in behavior categorizations that are less influenced by the initial comparison video.

In sum, the observed recategorization among no load participants in the current study might have been the result of direct adjustments to behavior categorizations or it might have been mediated by adjustments to expectancies. Both mechanisms would have the same end result in this paradigm (behavior recategorization). As a result, we are unable to definitively favor one mechanism over another. However, our own best guess is that perceivers adjust their expectancies in light of incoming performance information if they have the cognitive resources to do so. Future research will have to test this possibility directly.

The On-line Rating Methodology

Potential limitations. One could be concerned that our method changes the process by which participants perceive and rate target stimuli. Two types of interference seem possible: the on-line rating task could serve as a cognitive load, distracting participants from their primary task of perceiving the target, or the on-line rating task could force participants to evaluate the target differently than they otherwise would (i.e., engage in on-line vs. memory-based evaluation) (Hastie & Park, 1986).

For several reasons, we think the first type of interference is unlikely. First, the accuracy of video recall in the current study was very similar to that obtained in our other studies where the on-line rating procedure has not been employed (Weary, Reich, & Tobin, 2001; Weary, Tobin, & Reich, 2001). Second, if the on-line rating task effectively had put all participants under cognitive load, then we would not have observed the correction in on-line and postvideo ratings that we saw among participants in our no load condition. Finally, we went to great lengths to provide participants with adequate practice with the on-line rating procedure before they rated the main target: We first allowed them to practice right to left movements with the joystick and then had them rate how a target changed over time. This level of practice with the task should have freed up resources for other tasks, such as observing the child's performance and evaluating his intelligence.

Did the on-line rating task force participants into a different mode of evaluation? Did our procedure turn what typically is a memory-based evaluation into an on-line one? We also think this is unlikely. Given that participants in our study were given the goal of figuring out how intelligent the child was, they probably would have monitored the relevant performance information even in the absence of the joystick rating procedure. Moreover, researchers have argued that even without such explicit instructions, behavior categorizations and dispositional judgments typically are made on-line (Trope, 1989; Uleman, Newman, & Moskowitz, 1996). When using the on-line rating procedure in other areas, however, researchers should consider the issue of whether the process of interest is likely to be an on-line or memory-based one.

Future directions. Our findings, along with other recent research, illustrate the importance of looking at social judgments as they unfold over time (Kunda, Davies, Adams, & Spencer, 2002; Vallacher, Nowak, & Kaufman, 1994). Using this approach, we were able to examine judgmental influence and correction as they may occur in everyday life. In many different settings, behavior is continually monitored, such as in a job interview, in a classroom, or during a performance evaluation. Our on-line data suggest that perceptions of such behavior initially may be influenced by contextual factors, such as salient comparison standards (i.e., the previous job applicant, other students, other performers), regardless of one's level of cognitive distraction. However, these initial influences may be overcome as time goes on if perceivers have the necessary motivation and cognitive resources to adjust their inferences.

We think that the joystick-rating procedure used in the current study easily could be applied to other research questions that involve time-dependent changes in social cognitive processes, particularly those that involve judgmental corrections.[7] In the realm of dispositional inferences, for example, it may be interesting to examine directly and under various conditions when effortful Stage 2 corrections take place. In other areas of research, such as stereotyping and prejudice, researchers could use this methodology to examine whether perceivers who typically avoid making stereotypic judgments are initially as influenced by them, as are less motivated perceivers. Social cognitive research examining assimilation and contrast effects also could use an on-line rating procedure to examine when these effects occur, and whether the impact of contextual influences changes over time (i.e., early assimilation could turn into contrast if participants are aware of bias and overcorrect for it).

Conclusions

Although several researchers have suggested that some kinds of contrast in social judgments can occur automatically, only the most preliminary evidence of such effects exists in the literature (Moskowitz & Skurnik, 1999; Weary & Reich, 2001; Weary, Tobin, & Reich, 2001). The current research not only provides additional support for the efficiency of comparison-based contrast effects in dispositional judgments but illustrates over time and under optimal processing conditions the correction for such effects at the level of behavior categorizations. Indeed, the evidence provided by the current research allows us to confidently recommend that behavior recategorization be taken seriously as a potential means of correction of dispositional inferences. Although much future work will be necessary before we are able to specify precisely the behaviors for and the conditions under which such recategorization occurs, what is clear now is that parts of the dispositional inference process are more flexible than previously thought.

NOTES

1. Expectancies that were distinct, or separable, from the representation of the target's performance were found to exert a contrastive influence on target judgments, whereas those that were nondistinct exerted an assimilative influence (Weary, Tobin, & Reich, 2001).

2. Only one design was used in the unsuccessful comparison video because a failure on these designs typically took longer than a success.

3. All analyses in this article included gender as a covariate. Overall, very few gender effects were found. For this reason and because in no case did gender interact with any of our manipulations, we do not discuss such effects here.

4. This manipulation was intended to vary the distinctness of the comparison video. However, simply not rating the comparison video was insufficient to render it nondistinct from the main video because it was clearly a different child (and, hence, easily separable from the main child) and participants likely formed on-line impressions of the comparison performance even if they did not have to rate it. Because this manipulation proved unsuccessful, data are reported collapsed across this variable.

5. In a pilot study, in addition to manipulating participants' cognitive resources and the level of performance portrayed in the comparison video, we included a no-video control condition to examine whether the contrast we observed above was absolute contrast (i.e., relative to a group that received no comparison standard) or relative contrast. Specifically, 128 men ($N = 61$) and women ($N = 67$) were randomly assigned to conditions in this 2 (cognitive load: load, no load) × 3 (comparison video: successful, unsuccessful, none) between-participants factorial design. We examined their ratings during a period in which both load and no load participants in the comparison video conditions exhibited contrast in their on-line ratings: the last 20 s of the first and second video clips. We conducted a 3 (comparison video: unsuccessful, successful, none) × 2 (cognitive load: no load, load) × 2 (clip) mixed-model ANOVA with video clip (test item) serving as a repeated measure. In addition to a main effect of video clip, $F(1, 121) = 10.89$, $p < .01$, there also was a significant comparison video main effect, $F(2, 121) = 8.87$, $p < .001$. A Dunnett test revealed that on-line ratings made by both unsuccessful ($M = 6.36$) and successful ($M = 5.23$) comparison video groups differed from ratings made by control participants ($M = 5.87$), one-tailed ps = .046 and .013, respectively. Thus, the contrast we observed can be considered absolute, rather than just relative, contrast.

6. This mediational relationship holds up when comparison video condition is included as a predictor. In addition, although theoretically we would expect behavior categorizations to precede dispositional inferences, some may argue that support for the reverse model also could be found in our data. However, the reverse mediational model (dispositional inferences mediate the effects of early on-line impressions on later behavior categorizations) did not meet the necessary requirements. Specifically, among load participants, early on-line impressions still predicted final behavior categorizations when dispositional inferences were included in the model ($\beta = .58$, $t = 2.38$, $p < .05$), and among no load participants, later on-line impressions still were associated with final behavior categorizations when dispositional inferences were included in the model ($\beta = .45$, $t = 1.95$, $p = .06$). In addition, among no load participants, the drop in predictive power of later on-line impressions when dispositional inferences were included in the model was only marginal ($Z = 1.91$, $p = .06$).

7. Other methodologies have been used to identify temporal changes, including repeated assessments of stereotype activation at fixed intervals (Kunda, Davies, Adams, & Spencer, 2002), examination of cognitive and affective judgments (i.e., with affective judgments presumably tapping lingering influence that already has been removed from cognitive judgments) (Gilbert, Giesler, & Morris, 1995), and various on-line tracking methods. On-line tracking tasks have had participants move a pointer on a linear potentiometer (Gilbert & Osborne, 1989) or use a mouse to move a cursor toward or away from a target on a computer screen (Vallacher, Nowak, & Kaufman, 1994). Our joystick rating method was similar to these other on-line tracking methods but allowed our participants to rate behavior on a fixed dimension without having to divert their attention away from the stimulus material.

REFERENCES

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 253-285). New York: Cambridge University Press.

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.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189-211). New York: Guilford.

Gilbert, D. T., Giesler, R. B., & Morris, K. A. (1995). When comparisons arise. *Journal of Personality and Social Psychology*, *69*, 227-236.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*, 509-517.

Gilbert, D. T., Krull, D. S., & Pelham, B. W. (1988). Of thoughts unspoken: Social inference and the self-regulation of behavior. *Journal of Personality and Social Psychology*, *55*, 685-694.

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

Gilbert, D. T., & Osborne, R. E. (1989). Thinking backward: Some curable and incurable consequences of cognitive busyness. *Journal of Personality and Social Psychology*, *57*, 940-649.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*, 733-740.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, *93*, 258-268.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley.

Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper & Row.

Jarvis, W. B. G. (2000). MediaLab Research Software, Version 2000 [Computer software]. New York: Empirisoft. (Available at www.empirisoft.com)

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 220-265). New York: Academic Press.

Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151-174). Morristown, NJ: General Learning Press.

Kunda, Z. (1998). Parallel processing in person perception: Implications for two-stage models of attribution. In J. M. Darley & J. Cooper (Eds.), *Attribution and social interaction: The legacy of Edward E. Jones* (pp. 115-126). Washington, DC: American Psychological Association.

Kunda, Z., Davies, P. G., Adams, B. D., & Spencer, S. J. (2002). The dynamic time course of stereotype activation: Activation, dissipation, and resurrection. *Journal of Personality and Social Psychology*, *82*, 283-299.

Martin, L. L., & Seta, J. J. (1983). Perceptions of unity and distinctiveness as determinants of attraction. *Journal of Personality and Social Psychology*, *44*, 755-764.

Moskowitz, G. B., & Skurnik, I. W. (1999). Contrast effects as determined by the type of prime: Trait versus exemplar primes initiate processing strategies that differ in how accessible constructs are used. *Journal of Personality and Social Psychology*, *76*, 911-927.

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

Trope, Y. (1989). The multiple roles of context in dispositional judgment. In J. N. Bassili (Ed.), *On-line cognition in person perception* (pp. 67-97). Hillsdale, NJ: Lawrence Erlbaum.

Trope, Y. (1998). Dispositional bias in person perception: A hypothesis-testing perspective. In J. M. Darley & J. Cooper (Eds.), *Attribution and social interaction: The legacy of Edward E. Jones* (pp. 67-97). Washington, DC: American Psychological Association.

Trope, Y., & Alfieri, T. (1997). Effortfulness and flexibility of dispositional judgment processes. *Journal of Personality and Social Psychology*, *73*, 662-674.

Trope, Y., Cohen, O., & Alfieri, T. (1991). Behavior identification as a mediator of dispositional inference. *Journal of Personality and Social Psychology*, *61*, 873-883.

Trope, Y., Cohen, O., & Maoz, Y. (1988). The perceptual and inferential effects of situational inducements on dispositional attribution. *Journal of Personality and Social Psychology*, *55*, 165-177.

Trope, Y., & Gaunt, R. (1999). A dual-process model of overconfident attributional inferences. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 161-178). New York: Guilford.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology*, *28*, 211-279.

Vallacher, R. R., Nowak, A., & Kaufman, J. (1994). Intrinsic dynamics of social judgment. *Journal of Personality and Social Psychology*, *67*, 20-34.

Weary, G., & Reich, D. A. (2001). Attributional effects of conflicting chronic and temporary outcome expectancies: A case of automatic comparison and contrast. *Personality and Social Psychology Bulletin, 27*, 562-574.

Weary, G., Reich, D. A., & Tobin, S. J. (2001). The role of contextual constraints and chronic expectancies on behavior categorizations. *Personality and Social Psychology Bulletin*, *27*, 62-75.

Weary, G., Tobin, S. J., & Reich, D. A. (2001). Chronic and temporary distinct expectancies as comparison standards: Automatic contrast in dispositional judgments. *Journal of Personality and Social Psychology*, *80*, 365-380.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. *Advances in Experimental Social Psychology*, *29*, 141-208.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*, 117-142.

Received June 14, 2002

Revision accepted October 31, 2002