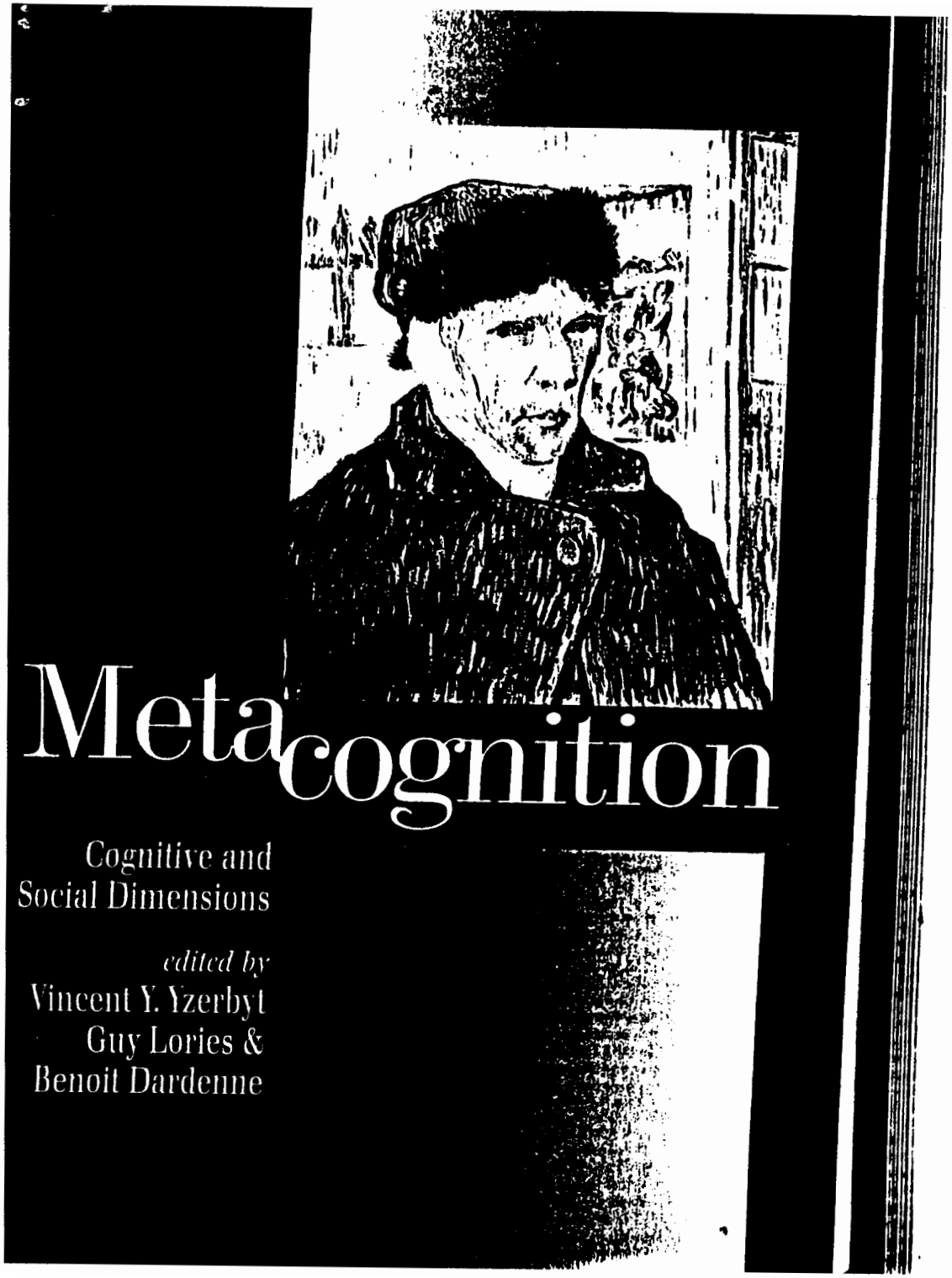

12

Correction and Metacognition: Are People Naive Dogmatists or Naive Empiricists during Social Judgments?

Leonard L. Martin and Diederik A. Stapel

Let's start with the obvious: (1) people can enact behaviors and (2) people can provide explanations for their behaviors. Now, the less obvious: What is the relation between the two? Do people consciously consider various features of the situation, weigh the pros and cons, develop a conscious intention, and then act upon that intention (cf. Fishbein & Ajzen, 1975)? Or do people initiate behaviors for reasons of which they may not be entirely aware, and only later attempt to piece together why they might have behaved as they did (cf. Bem, 1967; Nisbett & Wilson, 1977)?

In this chapter, we address this general issue in the area of social judgment. Obviously, people can make judgments and, obviously, people can provide rationales for their judgments, but what is the relation between the two? To paraphrase our opening question: Do people consciously consider various features of the judgment situation, weigh the positive and negative features of the context and target, develop a conscious understanding of the effects of these features, and then make judgments based upon that understanding? Or do people arrive at judgments for reasons of which they may not be entirely aware, and only later attempt to piece together why they might have made the judgments they did?

Each of these positions has its proponents in current social judgment theorizing. In this chapter, we compare and contrast these alternate views. First, we discuss the view that people's attempts to make their judgments accurate (i.e. remove bias from their judgments) are guided by their naive, metacognitive theories (Strack, 1992; Wegener & Petty, 1995; Wilson & Brekke, 1994). Then, we discuss the view that people's accuracy attempts are guided by non-conscious processes initiated by features of the general judgment setting (Martin & Achee, 1992). After presenting these two views, we discuss our own research on judgmental correction processes, and conclude from this work that people's conscious theories of social judgment are generally not causal, a priori, or accurate. Rather, these theories are descriptions of what people think they observed themselves doing while forming a judgment, and these theories influence judgments primarily when people are sensitized to (e.g. warned about) a particularly salient bias.

Metacognition

Cognitive and Social Dimensions

edited by Vincent Y. Yzerbyt,
Guy Lories and Benoit Dardenne

SAGE Publications
London • Thousand Oaks • New Delhi

12

Correction and Metacognition: Are People Naive Dogmatists or Naive Empiricists during Social Judgments?

Leonard L. Martin and Diederik A. Stapel

Let's start with the obvious: (1) people can enact behaviors and (2) people can provide explanations for their behaviors. Now, the less obvious: What is the relation between the two? Do people consciously consider various features of the situation, weigh the pros and cons, develop a conscious intention, and then act upon that intention (cf. Fishbein & Ajzen, 1975)? Or do people initiate behaviors for reasons of which they may not be entirely aware, and only later attempt to piece together why they might have behaved as they did (cf. Bem, 1967; Nisbett & Wilson, 1977)?

In this chapter, we address this general issue in the area of social judgment. Obviously, people can make judgments and, obviously, people can provide rationales for their judgments, but what is the relation between the two? To paraphrase our opening question: Do people consciously consider various features of the judgment situation, weigh the positive and negative features of the context and target, develop a conscious understanding of the effects of these features, and then make judgments based upon that understanding? Or do people arrive at judgments for reasons of which they may not be entirely aware, and only later attempt to piece together why they might have made the judgments they did?

Each of these positions has its proponents in current social judgment theorizing. In this chapter, we compare and contrast these alternate views. First, we discuss the view that people's attempts to make their judgments accurate (i.e. remove bias from their judgments) are guided by their naive, metacognitive theories (Strack, 1992; Wegener & Petty, 1995; Wilson & Brekke, 1994). Then, we discuss the view that people's accuracy attempts are guided by non-conscious processes initiated by features of the general judgment setting (Martin & Achee, 1992). After presenting these two views, we discuss our own research on judgmental correction processes, and conclude from this work that people's conscious theories of social judgment are generally not causal, a priori, or accurate. Rather, these theories are descriptions of what people think they observed themselves doing while forming a judgment, and these theories influence judgments primarily when people are sensitized to (e.g. warned about) a particularly salient bias.

The initial judgment models

It has generally been assumed that all judgments are relative to some context (see Eiser, 1990). We cannot describe an object as large or small, for example, without having some sort of standard in mind (i.e. large or small compared to what?). Consistent with this context-dependent view of judgment, a great deal of research has explored the effects of various contexts on people's judgments. This research can be divided roughly into two categories based on the nature of the contextual stimuli typically used. In research following the tradition of social judgment theory (Sherif & Hovland, 1961), the contextual stimuli have typically been exemplars that are obviously relevant to the target judgment, and that are judged along the same dimension as the target stimulus. For example, researchers in this tradition might ask participants to judge the size of a dog after having judged the size of an elephant or a mouse (cf. Herr, 1986; Herr, Sherman, & Fazio, 1983).

In priming research, on the other hand, the context has been construed not as another external stimulus to be judged, but as the cognitive concepts that are accessible to the judge at the time of the target judgment. The accessibility of these concepts has typically been manipulated in subtle, disguised ways (e.g. having participants read a series of words in what is ostensibly a color perception task in what is ostensibly an experiment unrelated to that in which the participants subsequently form their impression of the target). One consequence of these unobtrusive manipulations is that participants in priming studies typically do not realize that the contextual stimuli are in any way related to their target judgment (Martin, 1986; Stapel, Koomen, & Van der Pligt, 1997; Strack, Schwarz, Bless, Kuebler, & Waenke, 1993). In fact, priming can influence judgments even when participants are not aware of having been exposed to the contextual stimuli (Bargh & Pietromonaco, 1982; Winkielman & Schwarz, 1996).

Despite these (and other) differences (see Stapel et al., 1997), the social judgment research and the priming research make similar predictions regarding the conditions that lead to different types of context effects. Specifically, the initial theoretical models in both areas suggested that people assimilate their judgments of the target toward the implications of the context when certain context and target features overlap, but contrast their target judgments away from the implications of the context when the context and target do not show overlap (Herr et al., 1983).

The case for correction

The hypothesis that the degree of feature overlap between the context and target is crucial in determining the direction of context effects can take us far in understanding the effect of various contexts on judgments. Within the last 10 or so years, however, a number of studies have been published that

point to limitations on this hypothesis. These studies have shown that either assimilation or contrast can occur with a given degree of similarity between the target and the context (e.g. Martin, 1986; Martin, Seta, & Crelia, 1990; Strack et al., 1993). To account for such results, a number of investigators have begun to hypothesize that social judges may engage in more active processing than the initial models gave us reason to believe. Martin and Achee (1992), for example, suggested that a filter metaphor might be useful in explaining the occurrence of both assimilation and contrast to the same context-target configuration. The argument was that people have a wide range of information they could bring to bear on a judgment. This includes not only concepts or exemplars as suggested by the initial models, but also previously stored judgments, attitudes, moods, and emotions, non-mood bodily states such as arousal or confusion, scripts, communication rules, and more. For any given judgment, however, people bring only a subset of these into play. People may even fail to bring into play information that is highly accessible and applicable to the target. They may fail to do so because use of this information would be inappropriate in light of their current processing objectives. In short, the filter metaphor suggests that people are sensitive to the appropriateness of using a contextually induced response in addition to being sensitive to the response's similarity to the target response.

Conceptually similar positions have been proposed by others. Higgins (1989), for example, suggested that people judge the relevance of the context as well as its similarity to the target. According to Strack (1992), people assess neither relevance nor appropriateness; they assess representativeness. From this perspective, people do not use a contextually induced response in their target evaluation if that response does not appear to have been elicited by the target. Yzerbyt, Schadron, Leyens, and Rocher (1994) described a social judgeability model in which people do not use information in their judgments if they do not feel justified in doing so (e.g. characterizing someone in terms of a stereotype is unjustified). Wilson and Brekke (1994) suggested that people engage in correction processes in an attempt to avoid mental contamination, and Wegener and Petty (1995) suggested that people engage in correction processes in an attempt to avoid perceived bias.

Despite their differences, each of the models just described are based on the assumption that people are in some way actively influencing the effect of the context on their target judgment. Social judgments are not merely a function of the similarity between the target and the contextual stimuli. They are a function of the target, the contextual stimuli, the judge, and the judgment setting. One reasonable way to categorize these models is in terms of the amount of credit they accord to people's naive theories (i.e. people's verbal reports of the effect of the context on the target judgment). The most credit is accorded by Wegener and Petty (1995) who suggested not only that people's theories determine the direction and extent of their corrections but also that a consideration of people's naive theories may be useful in "organizing correction findings across different paradigms and domains"

(p. 39). Similar, though somewhat less expansive positions, have been taken by Wilson and Brekke (1994) and by Strack (1992).

Even less credit is accorded to people's naive theories by Martin and Achee (1992). Although these investigators described the possibility of both assimilation and contrast arising from people's use of theories (pp. 212–213), they qualified this possibility by suggesting that "the answer lies, in part, in people's theories" (p. 212). They suggested instead that most of the variance in correction is attributable to the general judgmental setting, or more precisely, to the implicit processing objectives activated by features of the setting. In discussing these models, we have divided them into two categories: Those that view correction as guided primarily by people's naive theories and those that view correction as guided primarily by features of the judgmental setting.

Theory-driven correction

There are at least three models that explicitly suggest that people's corrections for contextual influences are based primarily on people's naive theories (Strack, 1992; Wegener & Petty, 1995; Wilson & Brekke, 1994). Although these models differ from one another in certain respects, they share the following four assumptions: To correct for contextual influences, people must (1) be aware of being biased by the contextual stimuli; (2) be aware of the direction and magnitude of this bias; (3) be motivated to correct for this bias; and (4) have sufficient control over their responses to be able to correct. Unless all four of these conditions are satisfied, people will not correct. We present this general view by discussing one model in detail, namely Wegener and Petty's flexible correction model. The assumptions and predictions of this model are representative of the other models of theory-based correction, yet the flexible correction model has been empirically tested more than the other models. Consequently, it can be spoken about in more detail.

According to the flexible correction model (Petty & Wegener, 1994; Wegener & Petty, 1995), people possess a store of naive theories that specify the effects that various stimuli exert on various judgments. These theories are verbalizable, exist in a person's memory prior to the time of the target judgment, and play a causal role in the person's judgments. Before a research participant ever enters a psychology lab, for example, is placed in a positive mood, and asked to evaluate his or her life satisfaction, the participant could articulate the possible effect of his or her mood on his or her life satisfaction judgment. The participant may theorize, for example, that he or she will be predisposed to render a more favorable evaluation when in a positive as compared to a negative mood.

Of what use is this insight? Does it help people in their attempts to make their judgments more accurate? According to the flexible correction model, when people suspect that their target judgment may be in danger of being

biased, they retrieve their naive theories relevant to the judgment and the particular context in which they are making this judgment (e.g. a theory of the effects of mood on life satisfaction judgments). These theories allow people to determine the likely direction and extent of the effect of the context on their judgment. Armed with this information, people can attempt to remove the contextually induced bias from their judgment. They do this by adjusting their final target ratings in a direction opposite to the theorized bias and in an amount commensurate with the theorized amount of bias. Thus, people who suspect that a context might be exerting an assimilative effect on their ratings will adjust their ratings of the target in a direction away from the value of the context (i.e. contrast), whereas people who suspect that a context might be exerting a contrastive influence will adjust their ratings of the target back toward the context (i.e. assimilation).

It should be noted that the flexible correction model does not assume that people's theories are necessarily accurate. A person's theory might suggest assimilation, for example, when the context is actually biasing their judgments toward a contrast effect. In such cases, people's corrections follow their theories because corrections are aimed at removing the theorized bias, not the actual bias. As a result, "people may overcorrect, undercorrect, or even correct for a bias that does not exist" (Wegener & Petty, 1995, p. 38)

What are the conditions that make people access their theories and attempt to remove bias from their judgments? Unfortunately, the flexible correction model is mute on this point. The initial research has been concerned primarily with showing that when people correct for perceived bias, they do so in accordance with their naive theories. Thus, in the initial studies, participants were essentially instructed to correct (i.e. Don't let your rating of the context influence your rating of the targets). Wegener and Petty have suggested, however, that theory-based correction will only occur when people are willing and able to exert a sufficient amount of cognitive effort. This is because it is presumably more difficult to complete the correction process than it is to use one's initial reaction to the target as one's judgment (see Martin et al., 1990).

To summarize, the flexible correction model, like the other models of theory-based correction (Strack, 1992; Wilson & Brekke, 1994), is based on the assumption that people correct for unwanted influences on their judgments only when they are aware of the influence, aware of the direction and magnitude of the influence, and are motivated to engage in the correction process. Awareness of the direction and magnitude of the bias is assumed to come from people's naive theories which are verbalizable and in existence prior to the correction. Without these theories, people could be induced to correct but would not know exactly how to do so. With the theories, people at least have a direction in which to head. They may not succeed in their debiasing efforts, however, because their theories do not always accurately reflect the actual contextual influence.

Correction induced by the judgment setting

One alternative to the view that people's corrections are guided by a priori, verbalizable theories is the view that people's corrections are determined by processes that are beyond their introspective capabilities. We discuss this possibility in terms of Martin's set–reset model (Martin, 1985, 1986; Martin & Achee, 1992; Martin et al., 1990). Like the theory-based models of correction, the set–reset model assumes that people attempt to correct for biases in their target judgments and that this correction demands the expenditure of cognitive effort. The set–reset model also assumes that the direction and extent of some corrections can be influenced by people's naive theories (Martin & Achee, 1992, pp. 212–213). The model differs from the theory-based correction models, however, in according a much smaller role to these theories.

The overriding principle behind the set–reset model is that in making any judgment, people must (1) retrieve information that is pertinent to locating the target accurately on the dimension of judgment and (2) discard or fail to retrieve information that does not satisfy this objective. Although these processes may be triggered by conscious and verbalizable theories, it may often be the case that these processes are triggered relatively spontaneously by features of the judgment setting that may not be open to articulation. If we are asked to judge how tall a person is, for example, then it would be inappropriate for us to report that the person has blue eyes or is left-handed, because these features do not allow us to locate the target on the tallness dimension. It is unlikely, however, that we would be conscious of our decision to exclude eye color and handedness from our judgment. Phenomenologically, we make our judgments of tallness simply by noting how tall the person is. Similarly, if we are asked to judge our life as a whole, then it would be inappropriate for us to respond by considering only our social life. Although evaluation of our social life is relevant to judging our life as a whole, it is not the sole focus of a "life as a whole" judgment. But this reasoning may not come explicitly to mind as we make our life satisfaction judgment.

Correction in the absence of verbal reports of a contextual influence make sense if one keeps in mind the distinction between declarative and procedural knowledge (Anderson, 1982; Smith & Lerner, 1986; Gazzaniga, 1985). Roughly, declarative knowledge is knowledge of descriptive features, knowledge that can be verbalized. Procedural knowledge, on the other hand, is reflective of skills and abilities. It cannot be verbalized either because it has become automatized as a function of repeated practice or because of its inherent inaccessibility. By analogy, people may be able to declare that one object is farther away than another, but they cannot report on the extent to which they relied upon convergence, accommodation, and retinal disparity in making this judgment.

One useful characterization of procedural knowledge is in terms of implicit *if-then* rules or production systems (Anderson, 1982; Smith & Lerner, 1986). According to this characterization, if values of the stimulus

information match values in the *if* statement, then specified operations are initiated. Moreover, the assessment of the value match and the performance of the operations can take place beyond the person's awareness (Anderson, 1982; Smith & Lerner, 1986). Martin and Achee (1992) proposed that such a system might be involved in some corrections in social judgments. Their hypothetical reset production system was depicted as follows:

If my reaction to the target is <INAPPROPRIATE> then <RESET>.

The inappropriate slot is construed as an implicit filtering process that initiates a correction when it detects aspects of the target reaction that belong to something other than the target ("I like the target because I am in a good mood") or that are part of the target but a part that should not be included in the particular judgment the person is making ("I know the target is attractive but this has nothing to do with whether she is honest"). A number of factors that seem to satisfy this inappropriateness constraint have been identified. For example, people are unlikely to use contextual responses in forming their target judgments if the context is perceived as belonging to a different category than the target (Schwarz & Bless, 1992; Seta, Martin, & Capehart, 1979; Tajfel & Wilkes, 1963). People also fail to use contextual responses if those responses are attributed to a non-target source (Schwarz & Clore, 1983), or if the judge has adopted a communication rule that implicitly suggests that the contextual response be excluded from their judgment (Strack, Martin, & Schwarz, 1988). Explicit warnings or requests to exclude the contextual response also lead people to exclude that response (Leach, 1974; Stapel, Martin, & Schwarz, in press; Wegener & Petty, 1995). People are also less likely to incorporate their context response into their target judgment when they rate the context and the target in sequence rather than simultaneously (Martin & Seta, 1983; Byrne, Lamberth, Palmer, & London, 1969). And people attempt to avoid target-related reactions if those reactions violate social norms, as might be the case with stereotypes (Devine, 1989; Yzerbyt et al, 1994).

According to the set–reset model, once a reaction has been categorized as inappropriate, people may attempt to partial this reaction from their true reaction to the target (see Figure 12.1). In the context of the set–reset model, this type of correction is called resetting, and is thought to be accomplished by focusing on features of the target reaction that distinguish it from the context reaction, and then using these distinctive features to generate a new reaction. The result is a shift in the target judgment away from the implications of the contextually induced reaction. Because correction involves the disuse of the initial reaction and the generation of another, correction is assumed to demand the expenditure of more cognitive effort than does going with the initial reaction (Martin et al., 1990).

When the overlapping contextual reaction is not instantiated as inappropriate, people incorporate the reaction into their target judgment. The result is assimilation of the target judgment toward the contextually induced reaction.

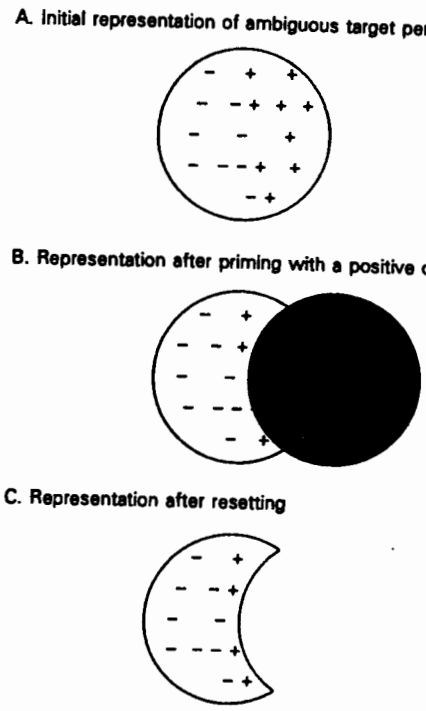

Figure 12.1 *Metaphorical representation of the set–reset process*

Dogmatists vs. empiricists

Both the theory-based models of correction as well as the set–reset model suggest that people try to place the target accurately on the dimension of judgment, and that they do this by removing what they perceive to be biasing contextual influences. The models differ, however, in their depiction of this debiasing process. According to the theory-based view, people have conscious theories of the direction and extent of the contextual bias, and use these theories to guide their corrections. It is in this sense that the theory-based correction models portray people as naive dogmatists. People have a priori theories of how the context influences their judgments and they use these theories to guide their corrections regardless of the accuracy of the theories.

The alternate view is that features of the judgment situation activate production systems which, in turn, initiate various judgment operations, including corrections, beyond the judge's awareness. Having observed themselves experiencing the conscious thoughts and feelings that are the

output of the production systems, however, people can make inferences about the probable influence of the contextual factors. In this way, people's theories and explanations come after the judgment to be explained (Bem, 1967; Griffin & Buehler, 1993). To the extent that this is true, people can be characterized as naive empiricists. They wait to see the conscious effects of various contextual factors before developing a theory to explain these effects.

Is there a way to tease these two views apart? In other words, can we tell if people act more like naive dogmatists or naive empiricists during social judgments? We think so.

Predicting the uncorrected reaction

One potentially interesting way in which the flexible correction model and the set–reset model differ from one another is in their treatment of uncorrected reactions (i.e. the effects of a context when judges do not correct for the contextual influence). The flexible correction model does not include factors that allow it to predict the uncorrected effect of any given context. From a flexible correction perspective, we can know the uncorrected effect only by exposing participants to a context and observing the resultant judgments when participants have not been induced to correct. A concrete example of this logic can be seen in a study by Wegener and Petty (1995). They began by providing participants with a series of context-target configurations and asked the participants to say what the effect of the context might be on the target. They found, for example, that most participants believed that their ratings of a product would be biased toward desirability if the product was endorsed by attractive as compared to unattractive women (i.e. an assimilation effect). On the other hand, most participants also believed that their ratings of moderately attractive women would be biased away from their ratings of extremely attractive or extremely unattractive women (i.e. a contrast effect).

After establishing that there were sets of stimuli for which participants held theories of either assimilation and contrast, Wegener and Petty (1995) had participants rate these stimuli. Half of the participants were asked, without further elaboration, to rate the context and target items. The remaining participants were given an explicit warning not to let their judgments of the context influence their judgments of the target. Although this warning informed participants of a possible bias, it did not specify the direction or magnitude of that bias. It was assumed that this information would be gleaned by participants from their naive theories. The results were consistent with this hypothesis.

When participants simply rated the context and target stimuli, their target judgments reflected assimilation when participants rated stimuli they had earlier theorized would lead to assimilation, but reflected contrast when they rated stimuli they had earlier theorized would lead to contrast. When participants had been instructed to remove the contextual bias, however, their judgments showed the opposite pattern. There was assimilation when participants rated stimuli they had earlier theorized would lead to contrast, but contrast when they rated stimuli they had earlier theorized would lead to assimilation.

According to Wegener and Petty (1995), when participants were not instructed to remove the contextual bias, their target judgments reflected the uncorrected effect of the context. When participants were instructed to correct, however, their target judgments reflected an effortful correction in a direction opposite to the contextual influence suggested by their theories. Thus, we can conclude that contrast is the uncorrected effect of rating moderately attractive women after rating either extremely attractive or extremely unattractive women because participants who rated these stimuli produced contrast when they had not been instructed to correct.

The set–reset model, on the other hand, suggests that aspects of the stimulus configuration can influence the uncorrected reaction. As can be seen in Figure 12.1, p. 235, the Venn diagram typically used to depict reset contrast depicts an overlap between the representation of the target and the representation of the context (e.g. priming with an applicable concept). This overlap is crucial because, according to the set–reset model, failure to remove this overlap results in assimilation, whereas reset contrast occurs only when people partial out this overlap. So, if we know the degree of overlap, we can predict the likely uncorrected effect. Assimilation will be the uncorrected effect when there is overlap between the context and the target, whereas contrast will be the uncorrected effect when there is no overlap between the context and the target. A consideration of these set–reset assumptions suggests an alternate interpretation of Wegener and Petty's (1995) results.

From a flexible correction perspective, we assume that contrast is the default effect of rating moderately attractive women in the context of either extremely attractive or extremely unattractive women because judgments of participants who rated such stimuli showed contrast. From a set–reset perspective, however, contrast is not the default with these stimuli. This is because with these stimuli, there is an overlap between the representations of the target and the context. Specifically, the extremely attractive women possessed attractive features, the extremely unattractive women possessed unattractive features, and the moderately attractive women possessed some attractive and some unattractive features. So, according to the set–reset model, there should be assimilation when participants do not correct for the contextual influence (i.e. partial out the overlap). If this is true, then why did Wegener and Petty (1995) obtain contrast with these stimuli when participants were not instructed to correct for the contextual bias?

Recall that the set–reset model allows for implicitly initiated correction (i.e. production systems). This means that participants need not be able to verbalize the contextual influence nor do they need to be aware of having engaged in correction. These assumptions leave open the possibility that participants may correct implicitly or spontaneously even when they have not received explicit instructions to correct. The result would be a contrast

effect that would appear to arise in the absence of a correction. It follows, therefore, that when participants are explicitly instructed to correct in such a case, their ratings would reflect assimilation because these participants would be correcting for the output of the earlier, implicit reset (i.e. a contrast effect).

Obviously, this interpretation is speculative. And, although it allows the set–reset model to account for the Wegener and Petty data, it also raises two questions. First, is it possible to find evidence of implicit resetting in the Wegener and Petty paradigm? Second, why would participants warned about a contextual bias correct if they had already corrected implicitly? We address each of these questions in turn.

Cognitive effort and implicit corrections

Both the flexible correction model and the set–reset model suggest that judgments rendered with little cognitive effort are likely to reflect the initial, uncorrected effect of a context. The models differ, however, on what they expect this effect to be when participants rate moderate stimuli in the context of more extreme stimuli (i.e. moderately attractive women in the context of either extremely attractive or extremely unattractive women). Wegener and Petty found that the uncorrected effect with such stimuli was contrast. According to the set–reset model the uncorrected effect with such stimuli is assimilation, but this assimilation should show up only when participants exert sufficiently little cognitive effort that they do not engage even in implicit correction.

To test these ideas, Martin (1997) had female participants rate the attractiveness of moderately attractive males in the context of either very attractive or very unattractive males (stimuli for which these participants held theories of contrast). Half of the participants were run under conditions similar to those of Wegener and Petty (1995). Specifically, some participants were asked merely to rate the contextual and target stimuli, whereas others, when asked to rate these stimuli, were instructed to keep their ratings of the context stimuli from influencing their ratings of the targets. Although the results in these conditions cannot tease the two models apart, they can tell us if we captured the same phenomena as Wegener and Petty. If we did, then we should see judgments of the unwarned participants reflecting contrast, but judgments of the warned participants reflecting assimilation.

It is the judgments of a second group of participants that are useful in testing the predictions about the uncorrected effect of the context. In this second group, all participants rated the contextual and target stimuli without receiving any mention of the contextual bias. Half of these participants, however, were induced to exert little cognitive effort in forming their judgments, whereas half were induced to exert considerable cognitive effort (Martin et al., 1990). Specifically, participants in the low effort condition were told that the experimenter was interested in "top of the head" judgments and that they were to exert little effort in forming their judgment, respond with the first thoughts that came to them, and put no name or other identifying mark on the paper. Participants in the high effort condition, on the other hand, were told to put their name on their papers and be as accurate as possible, so that later in the experiment they could justify their ratings.

According to the flexible correction model, when participants do not exert sufficient cognitive effort, their judgments should reflect the uncorrected effect of the context. Because contrast was reflected in the judgments of the unwarned participants in the Wegener and Petty study (and presumably what would be obtained in our parallel conditions), the flexible correction model leads us to conclude that judgments of the low effort participants will reflect contrast (i.e. the default, uncorrected effect).

The high effort participants, on the other hand, might exert sufficient effort to retrieve their theories and use them in correcting for the theorized bias. Because participants held theories of contrast for these stimuli, the corrections of these participants should produce assimilation, a shift away from the theorized bias. In sum, the flexible correction model leads us to expect that judgments of the low effort participants will reflect contrast, whereas judgments of the high effort participants will reflect assimilation.

The set–reset model, on the other hand, is based on the assumption that, with these stimuli, there is some overlap between the representations of the context and the target. Participants who do not exert much effort may not partial out this overlap, whereas participants who exert high effort may partial it out. Thus, the set–reset model predicts the mirror image of the flexible correction model: assimilation among the low effort participants, but contrast among the high effort participants.

As can be seen in Figure 12.2, the results supported the set–reset hypotheses. When participants did not exert cognitive effort, they rated the target as more attractive in the positive context than in the negative context (assimilation). When participants exerted effort, however, they rated the target as more attractive in the negative than in the positive context (contrast).

These results suggest that the judgments of Wegener and Petty's unwarned participants did not reflect the uncorrected effect of the context. Their unwarned participants, like ours, showed contrast in their judgments. Our low effort participants, on the other hand, showed assimilation. This pattern suggests that even though the unwarned participants were not alerted to a contextual bias, they nevertheless completed an implicit correction process. Our results fit with the flexible correction assumption that the uncorrected effect of a context can be observed when people do not exert sufficient cognitive effort, but they suggest that the no-warning condition used by Wegener and Petty did not produce a low enough level of cognitive effort to reveal the true uncorrected effect. Without a sufficiently low level of effort, participants (even unwarned ones) may complete an initial, implicit correction.

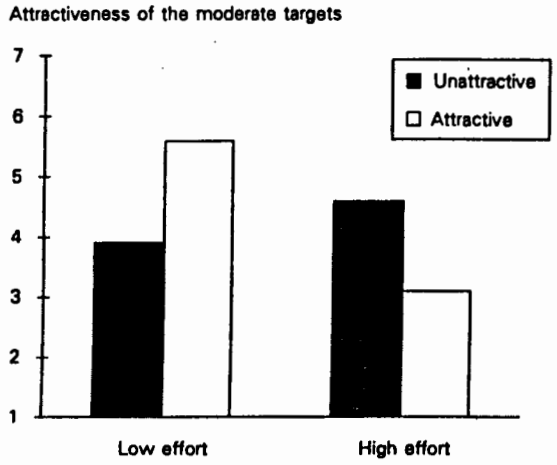

Figure 12.2 *Ratings of target persons as a function of level of effort and the attractiveness of the context persons*

What are the implications of the results in the high effort, accuracy condition? Participants given accuracy instructions clearly performed a reset. Their judgments reflected contrast, not the uncorrected assimilation seen among the low effort participants. This contrast was not a correction in the direction recommended by the participants' a priori, verbal theories, however. Participants had indicated in pilot testing that they thought the context would have a contrastive influence. So, if participants were correcting away from the bias suggested by their theories, then they would have corrected in the direction of assimilation. The accuracy participants, however, corrected away from the true uncorrected effect (low effort assimilation), toward the theorized uncorrected effect.

In fact, the only participants to correct in a direction recommended by the a priori theories were those given Wegener and Petty's blatant warning. It may be that the flexible correction process holds only when people are explicitly instructed to correct and are conscious of the output of their initial judgmental processes.

In sum, the data suggest that people act more like naive empiricists than naive dogmatists. Exposure to the context-target configurations seemed to engender an implicit correction process (i.e. resetting). Participants were unaware of this process, but were aware of its output. When asked what the effect of the context might be on their judgments (i.e. when asked their theory), they reported contrast, not the original uncorrected assimilation effect. Similarly, when given a blatant warning to correct, they corrected for contrast (i.e. the output of the initial correction process), not assimilation (i.e. the low effort default effect).

Correction or communication?

If the set–reset interpretation just described is correct, then the blatantly warned participants corrected for a judgment they had already corrected. Specifically, they performed an implicit correction followed by an explicit correction. Why would participants engage in this kind of double correction? There are at least two reasons why this might be the case. First, participants may have been unaware of having made the first correction. If this initial correction was indeed implicit, then participants would have been aware only of its output. This possibility is consistent with participants' self-reported theories. As noted earlier, when participants reported what the effect of the context would be, they reported contrast (i.e. the output of the initial, implicit reset), not the actual, uncorrected effect (i.e. assimilation). A second factor that might make people correct after having already corrected lies in the correction instructions themselves. Note that the blatant warning used by Martin and Wegener and Petty conveys a strong demand to correct. Following the rules of communication (Grice, 1975), participants may ask themselves "Why would the experimenter tell me to correct for the influence of the context unless there was in fact such an influence?" (see Leach, 1974; Schwarz, Strack, & Mai, 1991; Strack et al., 1988). In short, participants may have corrected twice because, in a sense, they thought they were being asked to do so.

Would participants show the same tendency to correct if the warning was less demanding? Suppose, for example, that participants were told to correct only if they actually detected a bias. With this warning, there is no implication that a bias exists, so there is no demand to correct. Thus, participants receiving a conditional warning might correct only if they actually detected a bias. This hypothesis can be tested by having participants receive either a blatant or a conditional warning while rating context-target configurations in which the biasing effect of the context is either subtle or obvious. If the demand hypothesis is valid, then conditionally warned participants may correct only when the bias is obvious, whereas blatantly warned participants may correct regardless of the level of bias.

This hypothesis was tested by Stapel, Martin, and Schwarz (in press). Following Wegener and Petty, they had participants rate exotic vacation spots (e.g. Hawaii) before rating the desirability of the weather in some less desirable midwestern US cities (e.g. Indianapolis). Some participants rated these vacation spots on the same dimension on which they subsequently rated the cities (i.e. desirability of weather). Other participants rated people's job satisfaction in the vacation spots, and then rated the desirability of the weather in the midwestern cities. It was assumed that the biasing effect of rating the vacation spots would be more apparent when participants judged the vacation spots and cities on the same dimension (cf. Brown, 1953).

Within each of these two conditions, some participants received the blatant warning used by Wegener and Petty (1995), whereas some received

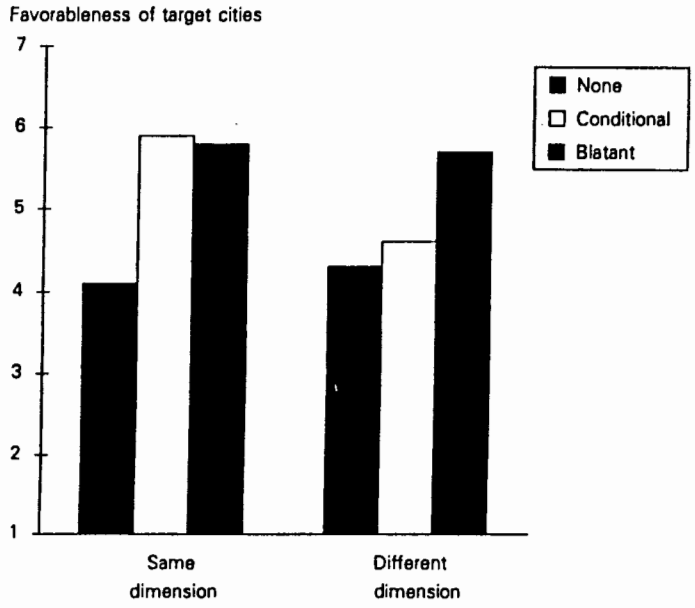

Figure 12.3 *Rating of target cities as a function of correction instruction and dimension of judgment*

a more subtle warning. The blatantly warned participants were instructed to keep their perceptions of the contextual stimuli from influencing their target judgments. The conditionally warned participants were instructed as follows: "Please try to make sure that your ratings of the desirability of the weather in the locations below reflect your true response. When you feel there is something that may have an unwanted influence on your ratings, please try to adjust for that influence." A third group of participants, the control group, received no warning at all.

If the blatant warning acts as a communication demand to alter one's target judgment, then we would expect to see participants given the blatant warning correcting regardless of the dimension on which they rated the cities and vacation spots. If the conditional warning, on the other hand, makes people sensitive to contextual influences, then we would expect to find that participants given the conditional warning would correct when they rated the cities and vacation spots on the same dimension, but not when they rated the two on different dimensions.

As can be seen in Figure 12.3, the results supported these hypotheses. Note, first of all, that participants in the control condition (i.e. those given no warning) did not correct for the implicit contrast. They judged the weather in the midwestern US cities to be undesirable relative to the vacation spots regardless of the dimension on which the vacation spots had been

judged. Participants asked to correct only if they detected a bias, however, corrected for the initial contrast when they rated the vacation spots and cities on the same dimension, but not when they rated these on different dimensions. Specifically, they judged the weather in the midwestern cities favorably after having rated the weather in the vacation spots, but rated it unfavorably after having rated people's job satisfaction in the vacation spots. Finally, participants who had received the blatant warning corrected for the initial contrast regardless of the dimension on which the vacation spots had been judged. Regardless of the context judgment, these participants rated the weather in the midwestern US cities as relatively desirable, an assimilative shift away from the initial contrastive effect of the context.

These results seem to suggest that the corrections obtained with the blatant warning used by Wegener and Petty (1995) do not reflect the operation of a spontaneous correction process. Rather, such corrections seem to reflect a compliance with communication demands. Participants correct because they think they are being asked to. This insight makes it easier to understand why participants might correct after having already made an implicit correction. They are essentially following the demands of the experimenter's communication. Unlike the conditionally warned participants, those receiving the blatant warning showed no sensitivity to the level of bias coming from the context.

Conclusions

In this chapter, we have explored the extent to which people are aware of what they are doing while making social judgments. Awareness, for us, referred to the extent to which people guide their judgment processes by use of verbal, a priori theories. Stated differently, can people give verbal accounts of the effects of various contexts on their judgments and do these verbal accounts subsequently influence their judgments? Our studies found little evidence that this was the case. Participants did not use their theories in the unwarned conditions of any of our studies, nor did they use their theories in either the low effort condition or the high effort condition in the Martin study.

The only participants whose judgments were suggestive of theory use were those who were either explicitly instructed to correct or who were conditionally warned in the presence of a blatant bias. The theory-based models of correction (see Wilson & Brekke, 1994 for a summary) tell us that people correct only when they are (1) aware of being biased, (2) aware of the direction and magnitude of this bias, (3) motivated to correct for this bias, and (4) have sufficient control over their responses to be able to correct.

The judgments of the conditionally warned participants fit nicely with these assumptions. These participants were instructed to correct only if they detected a bias, and they did, in fact, correct only when the influence of the

context was obvious. That these participants did not correct when the contextual stimuli were presented more subtly suggests that they did not detect a bias coming from the subtle context. In other words, they had no theory indicating that subtle presentation of the contextual stimuli could bias their judgments.

If the participants indeed had no theory of bias for the subtle context, then what induced the blatantly warned participants to correct for the subtly presented contextual stimuli? Our suggestion was that the blatant warning (i.e. don't let your ratings of the context influence your ratings of the target) induced compliance with a communication rule. The blatantly warned participants corrected because they were essentially being asked to do so.

Not only was correction by means of verbalizable, a priori theories rare in our studies, but we also found evidence that people can sometimes correct in a direction opposite to these theories. Specifically, the high effort participants in the Martin study rendered judgments reflective of contrast. This was clearly a correction relative to the assimilative default of the low effort participants. Because the participants' theories suggested that the context was exerting a contrastive bias, if participants had used these theories to guide their corrections, then they would have corrected in the direction of assimilation. The high accuracy participants, however, showed contrast relative to the low effort participants.

We should also note that to the extent that verbalizable theories came into play, they did not do so with an a priori status. The results suggest that participants developed their theories after engaging in an implicit correction process and reading off the output of this process. This is evidenced by the fact that participants' theories reflected contrast, not the default assimilative effect of the context. Similarly, when participants used their theories to guide their corrections, they corrected back in the direction of the initial default assimilation.

So, what implications do our results have for theory-based models of correction? Consistent with these models, we found that sensitizing participants to a bias can lead them to correct in a direction away from their verbal theories (the conditionally warned participants in the Stapel et al. studies). And we showed that people's theories are not always accurate. Participants had theorized that the effect of the context was contrast, whereas the actual default was assimilation.

Qualification of the theory-based view, on the other hand, came with the possibility that the blatant warning instructions often used in tests of theory-based models of correction may reflect simple compliance with communication demands. And there was at least one finding that seemed genuinely inconsistent with a theory-based view of correction. Participants in the high effort, accuracy condition of the Martin study corrected in a direction opposite to the a priori theories.

In sum, we see our results as more supportive of an empiricist view of social judgment than a dogmatist view. Exposure to the judgment situation (i.e. context, target, processing objectives) seemed to induce participants to perform certain judgment operations (e.g. resetting) and seemed to do so in a way that was beyond the participants' awareness. After observing the conscious output of these operations, however, participants were able to develop post hoc theories of how the contextual stimuli may have influenced them. Participants appeared to have done this, however, only when they received a warning to watch out for contextual biases.

At first blush, this empiricist view might seem to paint a rather negative portrait of social judges. One could conclude from our summary that social judges operate as mindless automatons, engaging in processing of which they are unaware, and rationalizing their actions after the fact. There is a more flattering interpretation, however. What the data indicate is that people's automatic processes are quite good at doing what they were developed to do. If people exert at least some minimal degree of cognitive effort, then their filtering processes can detect bias in their judgments and correct for that bias. Only after the filter has done its job does it provide people's conscious minds with the "unbiased" output. One implication of this arrangement is that people's verbal, conscious minds are free to pursue matters more interesting than scrutinizing each of their thoughts for bias. The general conclusion is that we may want to start placing more trust in our implicit processes. These processes may be more efficient than our verbal, a priori theories would lead us to believe.

References

Anderson, J. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406.

Bargh, J.A. & Pietromonaco, P. (1982). Automatic information processing and social perceptions: The influence of trait information presented outside of awareness. *Journal of Personality and Social Psychology, 43*, 437–449.

Bem, D.J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*, 183–200.

Brown, D.R. (1953). Stimulus-similarity and the anchoring of subjective scales. *The American Journal of Psychology, 66*, 199–214.

Byrne, D., Lamberth, J., Palmer, J., & London, O. (1969). Sequential effects as a function of explicit and implicit interpolated attraction responses. *Journal of Personality and Social Psychology, 13*, 70–78.

Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

Eiser, J.R. (1990). *Social judgment*. London: Open University Press.

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Gazzaniga, M. (1985). *The social brain*. New York: Basic Books.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.

Griffin, D. & Buehler, R. (1993). Role of construal processes in conformity and dissent. *Journal of Personality and Social Psychology, 65*, 657–669.

Herr, P.M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology, 51*, 1106–1115.

Assimilation and contrast effects. *Journal of Experimental Social Psychology, 19*, 323–340.
Higgins, E.T. (1989). Knowledge accessibility and activation: Subjectivity and suffering from unconscious sources. In J.S. Uleman & J.A. Bargh (Eds), *Unintended thought* (pp. 75–123). New York: Guilford Press.
Leach, C. (1974). The importance of instructions in assessing sequential effects in impression formation. *British Journal of Social and Clinical Psychology, 13*, 151–156.
Martin, L.L. (1985). *Categorization and differentiation: A set, re-set, comparison analysis of the effects of context on person perception.* New York: Springer-Verlag.
Martin, L.L. (1986). Set/reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology, 51*, 493–504.
Martin, L.L. (1997). Manuscript in preparation. University of Georgia.
Martin. L.L. & Achee, J.A. (1992). Beyond accessibility: The role of processing objectives in judgment. In L.L. Martin & A. Tesser (Eds), *The construction of social judgments* (pp. 195–216). Hillsdale, NJ: Lawrence Erlbaum Associates.
Martin, L.L. & Seta, J.J. (1983). Perceptions of unity and distinctiveness as determinants of attraction. *Journal of Personality and Social Psychology, 44*, 763–772.
Martin, L.L., Seta, J.J., & Crelia, R.A. (1990). Assimilation and contrast as a function of people's willingness to expend effort in forming an impression. *Journal of Personality and Social Psychology, 59*, 27–37.
Nisbett, R.E. & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.
Petty, R.E. & Wegener, D.T. (1994). Flexible correction processes in social judgment: Correcting for context-induced contrast. *Journal of Experimental Social Psychology, 29*, 137–165.
Schwarz, N. & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L.L. Martin & A. Tesser (Eds), *The construction of social judgments* (pp. 217–245). Hillsdale, NJ: Lawrence Erlbaum Associates.
Schwarz, N. & Clore, G.L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513–523.
Schwarz, N., Strack, F., & Mai, H.P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational analysis. *Public Opinion Quarterly, 55*, 3–23.
Seta, J.J., Martin, L.L., & Capehart, G. (1979). The effects of contrast and generalization on the attitude similarity-attraction relationship. *Journal of Personality and Social Psychology, 37*, 462–467.
Sherif, M. & Hovland, C.I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change.* New Haven, CT: Yale University Press.
Smith, E.R. & Lerner, M. (1986). Development and automatism of social judgments. *Journal of Personality and Social Psychology, 50*, 246–259.
Stapel, D.A., Koomen, W., & Van der Pligt, J. (1997). Categories of category accessibility: The impact of trait versus exemplar priming on person judgments. *Journal of Experimental Social Psychology, 33*, 44–76.
Stapel, D.A., Martin, L.L., & Schwarz, N. (in press). The smell of bias: What instigates correction processes in social judgments? *Personality and Social Psychology Bulletin.*
Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L.L. Martin & A. Tesser (Eds), *The construction of social judgments* (pp. 249–300). Hillsdale, NJ: Lawrence Erlbaum Associates.
Strack, F., Martin, L.L., & Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life-satisfaction. *European Journal of Social Psychology, 18*, 429–442.
Strack, F., Schwarz, N., Bless, H., Kuebler, A., & Waenke, M. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology, 23*, 53–62.
Psychology, 54, 101–114.
Wegener, D.T. & Petty, R.E. (1995). Flexible correction processes in social judgment: The role of naive theories of corrections for perceived bias. *Journal of Personality and Social Psychology, 68*, 36–51.
Wilson, T.D. & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*, 117–142.
Winkielman, P. & Schwarz, N. (1996). Contrast and assimilation with and without awareness: A test of the inclusion/exclusion model. Paper presented at the 68th annual meeting of the Midwestern Psychological Association, Chicago, IL, May.
Yzerbyt, V.Y., Schadron, G., Leyens, J.-P, & Rocher, S. (1994). Social judgeability: The impact of meta-informational cues on the use of stereotypes. *Journal of Personality and Social Psychology, 66*, 48–55.