THE FLEXIBLE CORRECTION MODEL: THE ROLE OF NAIVE THEORIES OF BIAS IN BIAS CORRECTION

Duane T. Wegener
Richard E. Petty

Life often presents us with situations in which it is important to assess the “true” qualities of a person or object, but in which some factor(s) might have affected (or might yet affect) our initial perceptions in an undesired way. For example, in the Reginald Denny case following the 1993 Los Angeles riots, jurors were asked to determine the guilt or innocence of two African-American defendants who were charged with violently assaulting a Caucasian truck driver. Some of the jurors in this case might have been likely to realize that in their culture many of the popular media portrayals of African-Americans are violent in nature. Yet, these jurors ideally would not want those portrayals to influence their perceptions of the particular defendants in the case. In fact, the justice system is based on the assumption that such portrayals will not influence jury verdicts.

In our work on bias correction, we have been struck by the variety of potentially biasing factors that can be identified—including situational influences such as media, social norms, and general culture, and personal influences such as transient mood states, motives (e.g., to manage impressions or agree with liked others), and salient beliefs—and we have been impressed by the apparent ubiquity of correction phenomena (which appear to span many areas of psychological inquiry). Yet, systematic investigations of bias correction are in their early stages. Although various researchers have discussed the notion of effortful cognitive processes overcoming initial (sometimes “automatic”) biases in a variety of settings (e.g., Brewer, 1988; Chaiken, Liberman, & Eagly, 1989; Devine, 1989; Kruglanski & Freund, 1983; Neuberg & Fiske, 1987; Petty & Cacioppo, 1986), little attention has been given, until recently, to the specific processes by which biases are overcome when effort is targeted toward “correction of bias.” That is, when
people identify potentially biasing factors, when and how do they attempt to remove bias from their perceptions and judgments?

In brief, we posit that corrections (i.e., attempts at preventing bias or debiasing initial perceptions of targets) are often the result of people consulting their naive theories of how potentially biasing factors might influence or have already influenced their perception of the target. As we explain shortly, this view differs from competing views of bias correction in that a view of corrections based on perceivers’ naive theories of bias allows for a more flexible set of corrections than those proposed by other current models of bias removal.

Correction phenomena have been discussed or studied in a variety of research areas including courtroom judgments, attribution, context and priming effects, and stereotyping and impression formation. In some of these areas (e.g., attribution; Gilbert, Pelham, & Krull, 1988), general discussions of correction have occurred, but little focus has been given to the specific processes by which corrections take place. Because much of the empirical work on bias correction has focused on the effects of target-irrelevant contexts on social judgments (e.g., Martin, 1986), and some explicit models outlining the underlying processes of correction have developed in this literature, our review of prior work begins there. Following our review of this work, we present our Flexible Correction Model (FCM) and empirical tests guided by this framework. The initial published tests of the FCM were conducted within a context effect/social judgment paradigm, but consistent with the generality of the model, we describe in this chapter a number of recently conducted studies in which the utility of the FCM is explored in a wider variety of domains (e.g., attribution, use of courtroom evidence, attitude change, etc.).

I. Ubiquity of Correction Phenomena and Past Discussions of Correction for Bias

People continually make assessments of the qualities of other people, places, objects, and issues in their environment. It is clear, however, that assessing the true qualities of a target—whether the target is a criminal defendant or the position taken by a political candidate—can often be a rather complex task. This is because perceptions of a target can be unduly influenced by a variety of personal and contextual factors present in the judgment situation (Petty & Wegener, 1993; Wegener & Petty, 1995b; Wilson & Brekke, 1994). Such undue influences might include situational or personal activation of concepts that influence interpretation of information
about the target (e.g., through semantic priming; Higgins, Rholes, & Jones, 1977; Martin, 1986; Srull & Wyer, 1980) or the activation of feelings that influence assessments of the target (e.g., Berkowitz & Troccoli, 1990; Forgas & Bower, 1987; Petty, Schumann, Richman, & Strathman, 1993; Schwarz & Clore, 1983; Wegener, Petty, & Klein, 1994). An individual might also learn that his or her initial perceptions of the target were based on incorrect, inappropriate, or incomplete information (e.g., Gilbert & Osborne, 1989; Golding, Fowler, Long, & Latta, 1990; W. Thompson, Fong, & Rosenhan, 1981; Wyer & Budesheim, 1987). Because of these potential biasing agents and others, accurate assessments of the target might often require some adjustment or “correction” to one’s initial reaction or some corrective action in the course of forming an initial reaction. Notably, attempts at correction do not necessarily render the assessments of targets more accurate. As we explain shortly, according to our flexible correction model, people might “overcorrect” (i.e., adjust judgments of the target farther than the biasing agent or agents had influenced assessments without the correction—leading to a bias in judgment opposite to the uncorrected bias), might correct for a perceived bias that does not exist, or might even correct in the wrong direction (exacerbating the initial bias).

When we began our correction work in 1990 (reported in Petty & Wegener, 1993; Wegener & Petty, 1992), correction phenomena were discussed in isolated research domains, and conceptual developments were aimed at explaining the phenomena in that domain. The work on attributional correction processes, discussed later in this chapter, is a good case in point (e.g., Gilbert et al., 1988; see also work on use versus disuse of information about targets in impression formation and jury settings; Golding et al., 1990; Thompson et al., 1981; Wyer & Budesheim, 1987). In the priming literature, however, broader models were developed to address biases and bias removal (e.g., Martin’s [1986] set/reset model). We first discuss the specific models of correction aimed initially at explaining assimilation and contrast effects resulting from priming. We next present our FCM and compare this perspective to the “partialling” (subtraction) models developed in the priming literature. Finally, we return to more domain-specific models of correction, such as those developed in the attribution literature and compare these positions with the flexible correction account.

A. PARTIALLLING MODELS OF CONTEXT CORRECTION: THE SET–RESET AND INCLUSION–EXCLUSION MODELS

Correction processes have recently played a prominent role in the area of context effects on impression formation and social judgment. As we
noted previously, in making judgments of targets, many kinds of factors are objectively not pertinent to the true qualities of targets (e.g., irrelevant mood states of the perceiver, unrelated information that serves to activate concepts in memory, etc.), but have been found to influence ratings of targets under certain conditions. In the context-effect literature, irrelevant contexts are discussed as either producing assimilation (i.e., making judgments of targets more like qualities of the context) or contrast (i.e., making judgments of targets less like qualities of the context).

Although many early discussions of assimilation and contrast focused on the distribution of contextual stimuli as responsible for these judgmental distortions (e.g., Helson, 1964; Parducci, 1965; Sherif & Hovland, 1961; see also Herr, Sherman, & Fazio, 1983), researchers have recently found evidence of assimilation and contrast effects even when the distribution of contextual stimuli has been held constant. For example, Strack, Schwarz, Bless, Kübler, and Wänke (1993) had research participants engage in a priming task that activated either positive or negative trait dimensions before they were asked to form an impression of an ambiguous target person (with a distraction task between the two activities). When Strack et al. (1993) reminded participants of the priming task before the impression formation task, they found contrast of target impressions away from the primes, but when participants were not reminded of the priming task, assimilation of impressions to the primes was found (see also Lombardi, Higgins, & Bargh, 1987; Martin, 1986; Newman & Uleman, 1990; Schwarz, Strack, & Mai, 1991).

Additional contexts and prompts to consider the context have shown similar results. For instance, Schwarz and Clore (1983) found that happy research participants rated their overall life satisfaction higher than sad participants when no attention was drawn to the source of the mood state. When attention was directed toward a potential cause of mood that was irrelevant to life satisfaction (e.g., the weather or experiment room), however, happy and sad participants rated their life satisfaction equally high. That is, when mood was made salient, and was made to appear irrelevant to the question at hand, participants adjusted their ratings of life satisfaction away from the initial mood-based responses; see Petty & Wegener, 1993; Strack & Hannover, 1996, for additional discussion). After discussing the "partialling" accounts of these effects, we describe the explanation provided by our FCM.¹

¹ Some researchers have attempted to account for such results using concepts other than correction per se. For example, attenuation of priming effects has also been discussed as due to extreme standards of comparison activated by the priming stimuli (Higgins, 1989), increased scrutiny of targets along the dimension of judgment activated by the priming event (Skowronski, Carlson, & Isham, 1993), or other "reprocessing" of target-relevant information (Ford & Kruglanski, 1995; E. Thompson, Roman, Moskowitz, Chaiken, & Bargh, 1994; see Wegener & Petty, 1995b, for further discussion).
B. THE SET–RESET MODEL

According to the set–reset model (Martin, 1986; Martin & Achee, 1992; Martin, Seta & Crelia, 1990), one's representation of an ambiguous target might include some positive and some negative elements (see top panel of Figure 1). When a context primes a set of thoughts (elements), some of the reactions to this context might overlap with the representation of the target—making the target seem more like the context (referred to as "setting;" see middle panel of Figure 1). When people realize that they are
thinking particular thoughts (in this case, positive thoughts) because they were reactions to the contextual stimuli, however, people will sometimes attempt to avoid using these thoughts in forming their impression of the target in an attempt to be accurate. In doing so, people attempt to “partial out” (subtract) the primed thoughts referred to as “resetting”). Because thoughts elicited by the context and thoughts elicited by the target are sometimes hard to distinguish, people might subtract out some of the elements of their true reaction to the target. Because of this, contrast away from the contextual stimuli can result (see bottom panel of Figure 1). Whereas setting leads to assimilation to the primed reaction, resetting might lead to contrast or to “judgments showing no effect of the prime or even a reduced assimilation effect” compared to conditions in which “setting” occurs (Martin & Achee, 1992, p. 212). Importantly, the partialling “corrections” that lead to these various outcomes are all in the same direction—away from initial reactions to the context.

A number of Martin’s studies have found very similar results using a wide variety of contextual stimuli and manipulations of reset likelihood. For instance, Martin (1986, Experiment 3) had research participants write self-relevant statements indicative of either a positive or negative mood. Although all research participants were stopped after writing four statements, some were led to believe that they would have to write eight statements rather than four (i.e., half of the participants believed that they were not finished with the writing task). Participants were then asked to form an impression of a target person described in terms that were ambiguous regarding the traits adventurous and reckless. Participants showed relative assimilation of their impressions to the mood statements when they believed that the writing task had been interrupted (i.e., they rated the target more positively after the positive than the negative writing task), but they showed relative contrast when they believed that the writing task had been finished (i.e., they rated the target more positively after the negative than the positive writing task).

Martin (1986) reasoned that people in the task-interrupted conditions should be more likely to continue to think about the writing task than those in the task-completed conditions (because of a “Zeigarnik” effect; Zeigarnik, 1927/1938). Because of this, people in the task-interrupted conditions would presumably find it more difficult to avoid the mood-activated thoughts in forming their impressions. In comparison, participants in the task-completed conditions would be more able to “partial out” the thoughts activated by the writing task.

Because resetting (correcting) includes an additional step beyond setting (i.e., resetting involves subtraction of elements activated by the context), Martin et al. (1990) reasoned that reset contrast effects require more cogni-
tive effort than assimilation effects. Using the same self-referent writing task as described earlier, Martin et al. (1990) found assimilation in three separate studies under conditions where past research had shown that cognitive effort was low, and found contrast under conditions where cognitive effort was expected to be high. That is, when research participants were distracted during the judgment task (Petty, Wells, & Brock, 1976), when participants' responses were not identifiable (Petty, Harkins, & Williams, 1980), and when participants were low in need for cognition (Cacioppo & Petty, 1982), assimilation was observed. Contrast was found in each of the studies when participants were not distracted, when their responses were individually identifiable, and when they were high in need for cognition, respectively.

C. THE INCLUSION–EXCLUSION MODEL

Schwarz and Bless (1992a,b) also proposed a “partialling” model that might account for such findings. The general tenets of the model can be summarized as follows: (a) contextual information that is recognized as irrelevant to the task is ignored and does not influence judgment (i.e., an “early exit” from the model occurs, and the contextual information is neither “included” in, nor “excluded” from, the person's representation of the target); (b) the default is to include potentially relevant information in the representation of the target—resulting in assimilation if no exclusion takes place; (c) features of the judgment task or communicative setting might trigger exclusion of potentially relevant contextual information; (d) excluding (subtracting) information from the representation of the target makes the target seem less like the contextual information, and excluded information might be used to set up an extreme judgmental standard with which the target can be compared (both of which lead to contrast effects); and (e) the emergence of contrast effects requires more processing steps, and more effort, than the emergence of assimilation effects (see Schwarz & Bless, 1992a, pp. 238–241).

Thus, similar to the set–reset model, the inclusion–exclusion model posits that the default (i.e., no-correction) bias associated with contexts is assimilation (because of including activated contextual information or reactions in the representation of the target). The inclusion–exclusion formulation is also in accord with the set–reset model when it posits that correction-producing factors in the judgment setting prompt exclusion (i.e., subtraction) of contextually activated information from the representation of the target. Thus, within both of these models, effortful corrections (i.e., resetting or exclusion) lead to target judgments less like the context than in no-
correction settings. Also, within both of these models, contrast effects are conceptualized as more effortful than assimilation effects (because the contrast effects are due to corrections, whereas the assimilation effects are not). The inclusion–exclusion model adds the feature that excluded (subtracted) information can then be used to construct a standard of comparison (which also requires cognitive effort and leads to contrast effects to the extent that the excluded information is extreme).

II. The Flexible Correction Model: Theory-Based Corrections

Partialling models of bias correction represented an important advance in determining when priming effects would or would not occur (or might even be reversed). However, some of the assumptions of these models (e.g., that contrast effects were more effortful than assimilation and that corrections always made assessments of targets less like reactions to the contextual stimuli) seemed to us unduly limiting if one attempted to apply them outside the priming-type of domain for which they were initially designed. Consider, for example, Helson’s (1964) work on adaptation level. In one set of studies, people placed their hands in buckets of cold (or hot) water. After a period of time, when the same hand was placed in a bucket of lukewarm water, the water was perceived as warmer (cooler) than would normally be the case. It seemed to us that such a contrast effect might likely occur without effortful corrections (because the processes are quite perceptual, and people are probably not aware of or concerned with “temperature” biases). However, if a person were to become aware of such a bias (e.g., because they placed both hands in the same water and realized that the “target” water cannot be two temperatures at once), corrections would likely make assessments of the water more (rather than less) like assessments of the contextual water (i.e., for the hand initially placed in the cold water, corrections make assessments of the target water cooler, but for the hand initially placed in the hot water, corrections make assessments of the target water warmer). Of course, such a correction cannot be explained by a partialling (subtraction) process that moves assessments of targets away from assessments of contextual stimuli.

An alternative way to organize context (and other) correction phenomena is to rely on a more flexible set of correction processes driven by respondents’ naive theories of how any given factor(s) have influenced their perceptions of the target. That is, unlike previous models of context correction that have focused on aspects such as the overlap between reactions to the context and target (Martin, 1986; Schwarz & Bless, 1992a) or changes in meaning of the scale anchors when the context is used to define
those anchors (Schwarz & Bless, 1992a, b), an alternative approach is to organize past correction work by focusing on the theories that respondents have or generate about how the context might have influenced (i.e., biased) their perceptions of the target and on the necessary steps for corrections to occur that are based on these naive theories of bias.

Basically, the FCM (Petty & Wegener, 1993; Wegener, 1994; Wegener & Petty, 1995b) holds that corrective action is aimed at removing (or avoiding) the bias that social perceivers believe is associated with the factor(s) at hand. Consider, for example, a situation in which a person is warned that a bias might be operating (e.g., in a jury setting in which evidence is deemed inadmissible). If the perceiver (juror) believes there is a large bias, he or she attempts to adjust assessments of the true qualities of the target more than if he or she believes that there is only a small (or no) bias. If the perceiver believes that the bias is to make the target seem higher on a dimension of judgment than would normally be the case, the perceiver attempts to adjust assessments of the target to be lower than his or her initial reaction toward the target; if the perceiver believes that the bias is to make the target seem lower on a dimension of judgment than would normally be the case, the perceiver attempts to adjust assessments of the target to be higher than his or her initial reaction toward the target. If the person is warned of a bias before actually encountering the target, the person might engage in theory-based corrections on-line as reactions to the target occur during exposure to the target. It is also possible that perceivers could use theories of bias to guide exposure choices so as to avoid perceived or expected biases (Wegener & Petty, 1995b; Wilson & Brekke, 1994; Wilson, Houston, & Meyers, in press). The greater the perceived bias, the more vigilant the perceiver would be in correcting on-line or in avoiding biasing information.

Work on theory-based correction follows directly from the notion that avoiding the influence of a stimulus would require awareness of the possible influence of the stimulus rather than awareness of the stimulus itself (see also Bargh, 1992; Higgins & Bargh, 1992; Jacoby & Kelley, 1990; Strack et al., 1993). For some time, researchers have noted that people are likely to possess or generate naive theories about how various contextual factors might influence of have influenced their perceptions of target objects (e.g., Nisbett & Wilson, 1977; Wilson, Laser, & Stone, 1982). For example, Nisbett and Wilson (1977) documented that people might believe that a factor influenced their perceptions (e.g., noise in an adjacent room) even if the factor had little demonstrable effect. As in this example, however, most of the earlier attention given to these naive theories was focused on the overall accuracy or inaccuracy of the theories (e.g., Nisbett & Wilson, 1977; Wilson et al., 1982) rather than on how those naive theories of bias might be used...
in attempts to remove biases from assessments of targets. In fact, as will be reviewed in the sections that follow, empirical evidence regarding the role of theories of bias in correction phenomena has only recently been reported (e.g., Petty & Wegener, 1993; Wegener & Petty, 1995b).

A. BASIC TENETS OF THE FLEXIBLE CORRECTION MODEL

As outlined in Table I, the FCM begins by noting that there is variation in the direction and magnitude of uncorrected effects of target-related, personal, and situational variables. That is, unlike the partialling models, which assume a default direction for context effects (i.e., “assimilation”) and other models discussed shortly that also assume a single default effect, such an assumption is explicitly rejected in the FCM framework. The FCM continues by noting that there are individual and situational differences in motivation and ability to identify potential sources of bias. If a person is unmotivated or unable to search for potential sources of bias, then his or her assessment of the qualities of the target will reflect his or her initial reaction to the target—whether this initial reaction is based on effortful scrutiny of the target or on more cursory analyses (see Petty & Cacioppo, 1986; Petty & Wegener, in press). If the perceiver is motivated and able to search for potential sources of bias, however, he or she will evaluate the potential biasing effect(s) of salient factors in the judgment setting (including factors external and internal to the perceiver). This is accomplished through accessing and/or generating naive theories of the biases associated with the salient factors. These perceptions of bias are naive theories in that a given perceiver is not likely to have direct access to the effect of the factor(s), nor is he or she likely to have the evidence necessary to definitively know the normative influence of the factor. Thus, the person's naive perception or theory of the potentially biasing effect of the factor is the person's best estimate of the effect of the factor, regardless of whether the perception is in any way accurate or not. The selection of a naive theory or the generation of one, however, can be influenced by one’s current reactions, to which the person does, of course, have direct access (e.g., “I'm reacting quite negatively toward the target, perhaps my reaction to the target is biased by the bad weather”). One’s reactions might be especially suggestive of a bias if they differ from an expected reaction.

If the person believes that no bias is operating, then the person’s assessment of the qualities of the target will reflect his or her initial reaction to the target. If, however, the perceiver believes that a bias is operating (regardless of whether this belief is incorrect or correct), and if the perceiver is motivated and able to correct (i.e., to attempt to “debias”) assessments
1. Across judgment targets, perceivers, and situations, there is variation in the direction and magnitude of default (i.e., "uncorrected") effects.

2. There are individual and situational differences in motivation and ability to identify potential sources of bias.
   a. If a person is unmotivated or unable to search for potential sources of bias, then his or her assessment of the qualities of the target will reflect his or her initial reaction to the target.
   b. To the extent that the perceiver is motivated and able to search for potential sources of bias, however, he or she will evaluate the potential biasing effect(s) of salient factors in the judgment setting (including factors external and internal to the perceiver). This is accomplished by consulting naive theories of the bias(es) associated with the salient factor(s).

3. If the person believes that no bias is operating, then the person’s assessment of the qualities of the target will reflect his or her initial reactions to the target. If the perceiver believes that a bias is operating (regardless of the accuracy or inaccuracy of this belief), then correction depends on the level of motivation and ability to engage in theory-based corrections.
   a. If the perceiver is unmotivated or unable to engage in corrections, then the person’s assessment of the target will reflect his or her initial reactions to the target.
   b. To the extent that the perceiver is motivated and able to correct (i.e., to attempt to "debias") assessments of the target, then the perceiver engages in a correction guided by the theory of bias.

4. Theory-guided corrections work in a direction opposite to the perceived bias and in a magnitude commensurate with the perceived magnitude of the bias. To the extent that the perceived bias is large, the theory of bias will create pressure toward greater adjustment (and/or vigilance in seeking qualities of the target that are consistent with greater adjustment) in assessments of the target.

5. Theory-based corrections can be undermined by a variety of factors that undermine the extent to which the theory of bias is viewed as applicable to the judgment target and setting, the extent to which the theory serves the perceiver’s judgment goals, and the extent to which the theory is accessible.

6. Although corrections generally require more motivation and ability (i.e., more cognitive effort) than lack of corrections (unless corrections become routinized), both corrected and uncorrected assessments of targets can vary in effort put into that assessment.

7. Just as differences in effort for uncorrected judgments create differences in persistence, resistance, etc., corrected assessments based on greater effort persist longer over time, are more resistant to attempts at changing assessments of the target, and are more likely to guide additional judgments and behavior toward the target than are corrected assessments based on lower levels of effort.

---

of the target, then the perceiver engages in a correction guided by the theory of bias. That is, the person attempts to adjust assessments of the target in a direction opposite to the perceived bias and in a magnitude
commensurate with the magnitude of the perceived bias. Of course, this effort does not take place separate from available information about the target. Part of this theory-guided correction can include seeking of information that would support these "corrected" assessments of the target. If no such supporting information can be found, this might be one factor that could undermine theory-based correction (we return later to additional factors that can undermine theory-based corrections and to potential theory-guided correction processes).

We assume that corrective processes ensue when people become aware of a potential bias and they are motivated and able to correct for that bias. Awareness of a possible bias can occur before, during, or after operation of a judgment process, however. Accordingly, corrections for bias need not occur only following initial reactions to the target, but people might also attempt to avoid an anticipated bias by changing how information about the judgment target is gathered, how the information is scrutinized, or by avoiding the biasing factor, if possible. We regard such attempts at avoidance of bias as "preemptive corrections," although common use of the term correction often refers to adjustment of existing reactions.

As noted earlier, within this model, no assumptions are made concerning the direction of the default (i.e., "uncorrected") effect of potentially biasing factors. For example, within the context-effect domain, a factor might make initial reactions toward the target more like reactions to the biasing context, make initial reactions less like reactions to the biasing context, or have no effect at all. Regardless of the default effect, corrections are driven by the perceptions of the bias. That is, corrections are aimed at removing perceived bias rather than actual bias. Although there are certainly cases in which perceived and actual bias coincide, the two elements are conceptually distinct. That is, a person might believe that a particular bias exists (and might attempt to remove that perceived bias) when no bias exists or even when a bias in the opposite direction is objectively present.2

A variety of factors might determine the nature of theories of bias and the likelihood that such theories guide attempts at correction. A theory of bias could be learned through experience and stored in memory, or could

---

2 Some of these notions are similar to the statements by Jacoby and Kelley (1987) that "to fully escape unconscious influences [by past events on current judgments] requires a theory that is sufficiently exact to satisfactorily replace subjective experience as a basis for judgment" (p. 333; italics added). According to the FCM, however, subjective experience and theory of influence can also be combined rather than discarding one's initial perception of the target and replacing it with a theory-based substitute (see Wegener & Petty, 1995b, for additional discussion). In fact, as we discuss in more detail subsequently, there might also be situations in which subjective experience is used in the formation of theories of bias, and subjective experience of the target might also provide input that moderates the perceived applicability of theories of bias to a given judgment target or situation.
be generated on-line to address a given judgment setting. In many settings, the theory of bias that is used might be some combination of a theory stored in memory along with adjustments to the theory based on the perceiver's experience of the judgment target and context. Although stored theories of bias are likely to have a greater base in past experience with the biasing agent or target, a theory of bias generated for the specific target and setting might be more likely to be viewed as applicable to the particular corrective attempt. The extent to which a theory of bias is viewed as applicable to a given setting and target would depend on factors such as the match between the theory and potentially biasing factor, the "strength" of the theory in terms of its integration with related knowledge structures and its accessibility in memory, the extent to which the perceiver experiences reactions to the target that are consistent with the theory of bias, and the "breadth" of the theory in terms of the situations and targets across which the theory is viewed as applicable. That is, if a theory of bias is perceived as applicable to the judgment setting and serves the judgment goals of the perceiver and is accessible when the perceiver assesses the target's qualities, that theory is likely to guide efforts at removing (or preventing) the perceived bias. To the extent that any or all of these properties are lacking, the theory becomes less likely to guide corrections, and the person might construct a new theory to account for this instance.

We generally assume that theory-based corrections require more effort than a lack of correction (at least when a potential bias is first encountered and considered), although at least some corrections might become routinized over many repetitions (cf. Smith, 1989; Smith, Stewart, & Buttram, 1992; see also Gilbert, McNulty, Giuliano, & Benson, 1992; Martin et al., 1990; Schwarz & Bless, 1992a). Importantly, however, we also assume that corrections can, themselves, vary in the extent of effort involved in the correction. Some theory-based corrections consist of a one-time adjustment to an assessment of the overall qualities of the target, whereas other theory-based corrections might also involve (re)interpretation of a great deal of information relevant to the target. Corrected assessments of targets that are based on a great deal of cognitive effort and that are well integrated with related knowledge structures are more likely to persist over time than are less well-integrated corrected assessments based on less cognitive effort (cf. Petty & Cacioppo, 1986; see later discussion of persistence and resis-

---

3 One might also regard the notion of "breadth" as the globality versus specificity of the theory of bias. That is, some biasing factors might be viewed by some people as only having effects on certain kinds of targets or in certain kinds of settings, whereas other factors (or the same factors considered by other people) might be viewed as having effects across many kinds of targets and/or settings.
tance to change of corrected assessments.\(^4\) In addition, uncorrected processes can vary in effort. That is, although corrections typically require some level of cognitive effort, effort per se does not necessarily imply that corrections will occur (or have occurred). These issues will be addressed in greater detail in later sections of the chapter.

Over the years, a number of researchers have made comments consistent with a theory-based view of corrections. Some earlier researchers speculated that respondents might use their naive perceptions of how a biasing factor has influenced them in order to adjust target ratings to compensate for the undue influence of the biasing factor (e.g., see Thompson et al., 1981; Wyer & Budesheim, 1987; for a more recent example, see Baumeister & Newman, 1994), but no general perspective was born out of these comments. Also, theoretical papers concurrent with the initial reports of research guided by the FCM (Wegener & Petty, 1992; Petty & Wegener, 1993) discussed the general possibility of theory-based corrections (see Strack, 1992a; Strack & Hannover, 1996).\(^5\) In a more recent review of mental contamination effects, Wilson and Brekke (1994) also noted that naive theories of bias might play a pivotal role in corrections. In addition, consistent with the flexible correction perspective (Petty & Wegener, 1993; Wegener & Petty, 1995b), Wilson and Brekke noted that people must be motivated to correct assessments of targets, must know the direction and magnitude of any bias at work, and must be able to adjust responses if successful debiasing is to occur. As in Nisbett and Wilson (1977) and Wilson et al. (1982), however, the focus of the Wilson and Brekke discussion was on the relative inability of people to accurately know which factors create biases (rather than providing any empirical evidence of particular means by which people attempt to correct assessments of targets).

In the following sections, we first relate the FCM to the literature existing when we began our research demonstrating the role of theories of bias in correction processes. Then, we review our initial work on theory-guided corrections.

\(^4\) Even though corrections are assumed to generally require more effort than a lack of corrections, this should not be taken as suggesting that "corrected" assessments of targets should necessarily persist over time or resist changes more than "uncorrected" assessments. Because "uncorrected" assessments are based more directly on reactions to the judgment target and such reactions might be recalled or reoccur upon additional presentations of the target, there might be a variety of settings in which "uncorrected" reactions tend to persist and resist change to a greater extent than "corrected" assessments (see Petty & Wegener, in press, for similar comments regarding changed and unchanged attitudes).

\(^5\) Strack (1992a) noted the possibility of corrections based on naive theories, but also retained assumptions of default outcomes of assimilation and increased effort for contrastive outcomes (see Strack, 1992a,b). Although such assumptions were consistent with the set–reset and inclusion–exclusion positions, we believe that they were inconsistent with the flexibility of theory-based corrections (see Wegener & Petty, 1995b). Even so, many of Strack's (1992a) statements regarding theory-based corrections are compatible with the FCM.
B. THE FLEXIBLE CORRECTION MODEL
AND THE EXISTING LITERATURE

1. Relation to the Existing Data

Is the flexible correction view that corrections are driven by naive theories of bias compatible with the existing literature? One characteristic of the studies explicitly investigating correction processes prior to development of the FCM was that corrections invariably moved judgments away from reactions to the biasing factor. That is, corrections made target ratings less like the content of priming episodes (e.g., Martin, 1986; Martin et al., 1990), less like inadmissible evidence (e.g., Thompson et al., 1981), less like contextual feeling states (Berkowitz & Troccoli, 1990; Schwarz & Clore, 1983), and less like the dispositional implications of behaviors (e.g., Gilbert et al., 1988, Osborne & Gilbert, 1992; see later section relating the FCM to attribution settings). It is noteworthy that within the context-effect literature, all of the studies used fairly moderate contextual stimuli that would normally be expected to lead to assimilation effects (e.g., Herr et al., 1983; Sherif & Hovland, 1961) when no correction processes are activated. Thus, to the extent that social perceivers believed that such factors produce biases in a direction consistent with reactions to the biasing factor, a theory-based correction would be entirely consistent with the observed results.

For example, consider the finding that presence of blatant primes (Martin, 1986; Martin et al., 1990) or of primes that have been made salient through reminding participants of their presence (Strack et al., 1993) can lead to contrast rather than assimilation. Rather than such effects being the result of a partialling or subtraction process (or of setting up a standard of comparison with the information subtracted from the representation of the target), it could be that people in those studies believed that assimilation was the likely uncorrected effect of the (blatant) priming contexts, and that corrections driven by these perceptions of bias were responsible for the observed effects. To the extent that people are found to believe that priming episodes create assimilative biases, then, past findings guided by partialling models of corrections might actually have been due to corrections based on perceivers’ naive theories of bias.

2. How Does the Theory-Based Perspective Differ from Previously Dominant Models?

In addition to providing a potential explanation of past correction effects, the FCM’s reliance on theory-based corrections makes a number of important predictions that differ from those made by previously dominant models. In this section, we focus on the differences between the FCM and the
context-effect (partialling) models of correction (i.e., the set-reset and inclusion-exclusion models). Later, we return to other areas in which general discussions of corrections have appeared (e.g., attribution), and we discuss research directions in those and other areas that are generated by the theory-based correction perspective.

As noted earlier, the first way in which the FCM differs from the partialling models regards assumptions about the direction of default (i.e., no-correction) processes. Whereas the set-reset and inclusion-exclusion models assume that default processes make reactions toward targets more like reactions toward the biasing factor (termed assimilation in that literature), the FCM assumes that default (uncorrected) processes can make initial reactions toward the target either more or less like reactions to the biasing factor. That is, using the context-effect terminology, default effects can be assimilation or contrast (or no effect). Regardless of the actual uncorrected effect, however, perceivers can believe that the bias is to make reactions to the target too much like reactions to the context (i.e., assimilation) or too little like the context (i.e., contrast). Perceivers can also believe that no bias occurred. Corrections are driven by these naive theories of how perceptions of the target have been (or might be) affected.

This leads to a second major difference between the models. That is, because theories of bias can be of either direction (toward or away from the biasing factor) corrections can flexibly go in either direction because they are guided by the theory of bias. An important implication of this view for the context-effect literature is that there can be multiple ways to arrive at assimilation and contrast effects. Whereas the set-reset and inclusion-exclusion models posit that assimilation is the default and contrast is the result of effortful correction, the FCM suggests that either effect could be the default or the result of corrections. To the extent that correction processes typically require cognitive effort (Gilbert et al., 1992; Martin et al., 1990; Schwarz & Bless, 1992a), correction-based contrast might require more effort than default (no-correction) assimilation; and correction-based assimilation might require more effort than default (no-correction) contrast (see Petty & Wegener, 1993; Wegener & Petty, 1995b). Thus, when one begins to look at corrections as driven by naive theories of bias, effort and judgment outcome become uncoupled. In comparison, according to a partialling view "the emergence of contrast effects requires extra processing steps, and more effort, than the emergence of assimilation effects" (Schwarz & Bless, 1992a, p. 240). In fact, according to a theory-based correction view, effort per se (even effort aimed at accuracy) does not necessarily mean that corrections will occur. According to a theory-based view, the perceiver must believe that a potential bias is operating for corrections to occur. Even if extensive processing of judgment-relevant informa-
tion occurs, if there is no identification or perception of bias, it is unlikely that corrections will take place.

For a comparison between theory-based corrections and partialling (subtraction) views, the case of corrections driven by a theory of default contrast is a crucial case. For example, a reset-based correction (in which contextually activated reactions are suppressed or "partialled out" of reactions to the target, e.g., Martin, 1986) cannot account for a correction that results in target ratings becoming more similar to reactions to the context. This is because the term set refers to "the use of a contextually activated response during formation of the target impression," and "when the contextual response is brought to bear [in forming the target impression], the evaluation of the target is assimilated toward the implications of the contextual stimuli." In addition, the term reset "refers to the suppressed use of the contextually activated response and the generation of a context-distinct response for the target." In resetting, "when the contextual response is not brought to bear [in forming the target impression], and the individual generates a context-distinct response, the evaluation of the target is contrasted with the implications of the contextual stimuli" (Martin, 1986, p. 495). Thus, because people "cannot use their reaction to the priming stimuli if their objective is to give their reaction to the target" (Martin & Achee, 1992, p. 210), reset-based corrections must result in reactions to the target being less like reactions to the context than when setting occurs (see Figure 1).

Therefore, in our initial tests of the FCM, we focused on aspects of the model that would distinguish it from partialling models of bias correction. However, as we have noted throughout our work on this topic, an important aspect of the FCM is the general applicability of theory-based correction notions across many phenomena and settings (e.g., see Petty & Wegener, 1993, pp. 161–163; Wegener & Petty, 1995b, pp. 48–49). Because of this, much of our ongoing work involves examination of flexible-correction hypotheses in domains as varied as persuasion, stereotyping, jury decision making, and attribution. Following our description of initial tests of the FCM, we discuss more recent work that demonstrates the generality of theory-based correction notions.

III. Initial Tests of the Flexible Correction Model

As we just noted, unlike the prior models of context correction, the FCM holds that contrast can sometimes be the uncorrected outcome in studies of social judgment, and effortful correction processes can move judgments
closer to the context rather than further away from it (i.e., if perceivers believe that contrast was the default bias). In fact, corrections for contrast constitute a critical test of the FCM in that partialling models only predict corrections that move ratings of targets away from the context. Thus, two of the prominent research questions in early research testing the FCM were (a) do people sometimes believe that contrast—rather than assimilation—is the result of uncorrected exposure to contexts?, and (b) if people believe that contrast is the uncorrected bias, do they correct assessments of targets by adjusting ratings to be more like the context than in “uncorrected” settings? Additional early research questions addressed the validity and utility of conclusions based on a theory-based view of corrections. The following sections describe the initial tests of the FCM.

A. THEORY IDENTIFICATION: DO PEOPLE SOMETIMES BELIEVE THAT CONTRAST RATHER THAN ASSIMILATION IS THE RESULT OF UNCORRECTED EXPOSURE TO CONTEXTUAL STIMULI?

Are people aware of contextual configurations that elicit contrast effects, or do people tend only to hold naive theories of assimilation for the effects of contextual stimuli? We reasoned that identification of contexts for which people believe uncorrected contrast effects are likely would enable us to demonstrate that people are capable of correcting by adjusting ratings of targets toward rather than away from assessments of the context. Because such corrections cannot be accounted for by traditional partialling models, identifying such a correction (and showing that it occurs when a theory of uncorrected contrast is held) would provide initial evidence that a theory-based correction model can go beyond partialling models of correction. Also, by showing that contexts such as those used in past studies of assimilation and contrast effects were seen by participants as likely to lead to uncorrected assimilation effects, it becomes possible that past results are a special case of more general theory-based corrections.

Of course, the first step in such an investigation is to identify contexts and judgments for which people believed the uncorrected effect to be contrast rather than assimilation. In order to examine this, we described various contexts and judgments to research participants and asked them what the natural effect of these contexts would be on people’s target ratings (Petty & Wegener, 1993; Study 1). In this study, all research participants received a questionnaire describing a number of contexts and judgments to be made. Two of the items described extreme contexts
that prior theory and research had suggested would bring about contrast effects (see Herr et al., 1983; Sherif & Hovland, 1961). If people ever believe that contrast is the uncorrected effect of a context, such beliefs might be most likely to be associated with contexts that have actually created contrast effects in past research (although people could also believe that contrast effects are created by contexts that do not actually bring about the effect). The two contrast contexts we selected were imagining being in the location of one's dream vacation before judging average locations (Dermer, Cohen, Jacobsen, & Anderson, 1979) and seeing a group of very attractive people before judging average-looking people (Kenrick & Gutierres, 1980). Two other items described contexts that would be expected to bring about assimilation effects if no correction processes were activated. These contexts included the effects of the priming of a trait on judgments of a person engaging in ambiguous behavior (Martin, 1986; Wyer & Srull, 1979) and the effects of positive mood on judgments of the pleasantness of everyday activities (Forgas & Moylan, 1987; Schwarz & Clore, 1983).

Importantly, people were found to view at least some stimuli as likely to bring about uncorrected contrast rather than assimilation effects. That is, respondents realized that being in the location of one's dream vacation would make average locations seem worse than would normally be the case ($p < .007$), and seeing a group of very attractive people would make average-looking people seem less attractive than would normally be the case ($p < .001$). Also, the kinds of stimuli used in past studies of context-effect correction processes were found to be associated with theories of uncorrected assimilation. That is, people believed that unscrambling sentences with hostile content would make a person described by ambiguous behaviors seem more hostile ($p < .007$), and that being in a good mood would make everyday activities seem more pleasant than would normally be the case ($p < .0001$).

Thus, it seemed possible to construct situations such that people perceive a contrast effect to be the natural biasing consequence of the context. That is, our theory-identification study demonstrated for the first time that individuals hold naive theories of both assimilation and contrast as the effects of contexts on target ratings. This implied that people believed that contrast might occur in the absence of correction processes, and more importantly that correction brought to bear on these judgments might be in a direction opposed to contrast rather than toward it. Of course, simply showing that participants believed that some contexts lead to contrast effects did not show that people knew how to correct and actually could correct for these expected contextual effects.
B. DO PEOPLE CORRECT FOR PERCEIVED CONTRAST IN A MANNER CONSISTENT WITH THEORY-GUIDED CORRECTIONS OR IN A MANNER CONSISTENT WITH PARTIALLING OF OVERLAP?

We expected that people induced to correct for a context they believed would create a contrast effect would correct for that context by adjusting their ratings of the target toward the context. In order to show that people can correct toward the context rather than away from it, we explicitly asked research participants in some conditions to attempt to keep contextual stimuli from influencing their judgments. Whereas past studies investigating correction processes used stimuli that would be expected to produce assimilation in the absence of correction processes, we used stimuli that would be more likely to produce contrast in the absence of correction processes.

In our first correction study (Petty & Wegener, 1993, Study 2), the context consisted of either five very popular and exciting vacation locations (e.g., Hawaii, Paris—the Extreme Positive context—an operationalization of the dream vacation context described in the theory-identification study) or a set of neutral American cities (e.g., Minneapolis, Houston—the Neutral context). After rating how much they would like to spend two weeks in each of the context locations, research participants either immediately rated two neutral target locations (Indianapolis and Kansas City) on the same scale (No Correction Instruction condition) or were first asked not to let their perceptions of the initial locations influence their perceptions of the targets (Correction Instruction condition). Because respondents believed that our extreme context would lead to contrast effects in the theory-identification study, we expected that asking people to avoid being influenced by the context would lead to a decrease in the contrast effect when compared with the no-instruction condition (i.e., participants would correct their judgments by adjusting their ratings toward the context). If, however, corrections for context always occur through processes in which judges subtract out their reactions to the context (e.g., Martin, 1986), then asking people who encounter an extreme context to avoid being influenced by their perceptions of the initial locations should either result in adjustments in judgments away from the context (if any overlap is perceived between reactions to the context and reactions to the target) or in no adjustments (if no overlap is perceived).

Ratings of the two target locations showed a significant Context X Correction Instruction interaction (see Figure 2). When no correction instruction was given, the targets were viewed more positively by research participants
who received the neutral context than by participants who received the extreme positive context (a contrast effect). When participants were asked not to be influenced by perceptions of the contextual locations, however, there was no significant difference in rating of the target between the neutral and positive context groups. In fact, the difference in means was actually in the direction of an assimilation effect. Importantly, the presence of the correction instruction led participants who had received an extreme context to rate the target locations significantly more positively compared to those who were not so instructed. That is, people corrected by rating the targets as more similar to the extreme positive context. Participants who received the neutral context showed no correction.

Thus, the results of this study indicated that people do not always make corrections by moving their judgments away from their reactions to the context. Whereas past studies of context effects found corrections leading to judgments of targets as less similar to the context (e.g., Martin, 1986; Martin et al., 1990; Schwarz & Clore, 1983), our study showed correction processes leading to judgments of targets as more similar to the context. Presumably, if participants in our study had believed that holding positive
views of the extreme contextual locations would bias them toward too positive a view of the target locations, their corrections would have been in the direction of further contrast. Instead, asking participants to correct for any influence of the contextual stimuli actually decreased (and nonsignificantly reversed) the contrast effect observed in the no-instruction conditions.

C. ARE CORRECTIONS FOR CONTRAST SIMPLY DUE TO RESPONSE LANGUAGE EFFECTS ACROSS CONDITIONS?

Although the results of our initial experiment appeared to unambiguously demonstrate adjustments in target assessments that were opposite in direction to those predicted by partialling models of bias correction, a relevant question is whether such adjustments represented changes in the assessments of targets per se, or simply changes in the way respondents used the scales to report an unchanging assessment. That is, if research participants used the extremely positive contextual items (e.g., Hawaii) to define the endpoint of the response scale (i.e., “like very much”), the meaning of this endpoint might have been more positive in the Extremely Positive than in the Neutral context conditions. If the scale was expanded in a positive direction when the context was very positive, then ratings of the target stimuli might be lower than in neutral conditions because of the expanded range of possible responses considered by respondents rather than any change in perceptions of the qualities of the target (see Ostrom & Upshaw, 1968). When correction instructions were given, however, participants might have taken such instructions to mean that the contextual stimuli should not be used in defining the scale endpoints. Thus, participants might have reverted to their “default” notions of what the scales meant yielding the “corrected” ratings of the target cities. If this explanation were true, then participants did not correct their perceptions of the target cities, but simply adjusted the end anchors of the rating scale.

Although there were many reasons to believe that corrections in our initial study were not due to scale anchoring (response language) effects (see Petty & Wegener, 1993, pp. 147–148), we thought it best to deal with such a possibility empirically. If the effects in our initial study were due to changes in the implied definitions of scale endpoints, then asking people not to be influenced by perceptions of initial locations should produce little change in target ratings if the definitions of response scale endpoints remain constant. That is, if scale endpoints include a specific referent that does not change across conditions, then corrections due to changes in scale meaning should be minimized. In the past, investigators attempting to
distinguish between response language and perceptual change effects have used response scales that are less likely to be redefined because of the existence of an external reference point—such as measuring distance in inches (Krantz & Campbell, 1961) or weight in ounces (Harvey & Campbell, 1963) rather than using the more subjective terms of “short/long” or “light/heavy” (cf. Biernat, Manis, & Nelson, 1991). Because liking has no “concrete” externally defined unit of measure, we used a different means to achieve a similar conceptual purpose. We used an explicitly defined subjective reference point that remained constant across conditions. That is, in some conditions, the response scale endpoints were labeled as representing participants' perception of a specific pair of extreme exemplars on the dimension of judgment.

In this study (Petty & Wegener, 1993, Study 3), all participants received the positive contextual items from the study described earlier. The scale anchors were either the same anchors used in the first correction study (Abstract condition—e.g., “like very much”) or included reference to perceptions of a pair of specific locations (Specific condition—“like as much as staying in Hawaii or Paris for 2 weeks [like very much]”). Respondents either rated the target locations immediately following the contextual locations or received the correction instructions before rating the targets. If the effects in our initial study were due solely to response language, a Correction Instruction X Scale Anchor interaction would occur—replicating our first study when scale anchors were abstract, but showing little effect of the correction instruction when scale anchors include specific referents. However, if our correction results were not due to response language, only a main effect of Correction Instruction would be found. As expected, only the main effect of Correction Instruction was significant ($p < .003$; Instruction X Scale Anchor interaction, $F < 1$). That is, regardless of whether the scale endpoints included reference to particular locations, the target items were rated more positively after correction instructions than when no instruction was given.

D. DO CORRECTIONS FOR CONTRAST OCCUR WITH MORE SUBTLE CUES TO REMOVE BIAS?

Past research had often used more subtle means of cuing correction processes than simply asking people to correct. For example, Schwarz and Clore (1983) merely asked respondents about the weather without telling them that the weather could have influenced judgments of life satisfaction. In the Martin (1986) studies, making participants believe that they were finished with the contextual rating task brought about corrections. Similarly,
asking target questions in a way that makes them appear to be asking for different information (Strack, Martin, & Schwarz, 1988) can bring about corrections in judgment. We have also used a variant of these procedures to explore the ability of subtle manipulations to elicit correction for contrast.

For example, in Petty and Wegener (1993; Study 4), some respondents rated contextual locations on a first page and then rated target locations on a second page in much the same way as in the no-correction conditions of our first correction study described previously (No Correction Cue condition). Other respondents rated the contextual locations on the first page, but on the top of the second page were told that for the "next group" of ratings there were "more vacation spots to consider." This instruction formed a sort of minimum baseline for correction in that the target locations were set aside as a "next group" but were a part of the same questionnaire from the same experimenter. This was a very subtle way of "ending" the first task (the context ratings)—a procedure found in past research to induce reset (correction) processes (see Martin, 1986).

Similar to the prior research using an explicit correction instruction, our more subtle correction cue led to a significant increase in the positivity of target ratings following the Extreme Positive context, but produced no changes in the ratings of targets following a Neutral context (p < .0004 for the Context × Correction Cue interaction; see Figure 3). Thus, it appears that at least some corrections toward the context occur even when cues to elicit such corrections are quite subtle. Therefore, even though much can still be learned through investigation of processes that take place when explicit instructions instigate correction processes, one goal for future research would be to identify additional factors and situations that more subtly foster corrections.

E. CAN THEORY-BASED CORRECTIONS PRODUCE OPPOSITE CORRECTIONS CORRESPONDING TO OPPOSITE THEORIES OF BIAS?

According to the FCM, corrections guided by judges' naive theories of bias go in opposite directions to the extent that the judges' theories of bias denote opposite biases. One way this could occur is if different people hold or generate different theories about the same context. For example, one person might think that blatantly hostile primes make target judgments more hostile, but another person might think that blatantly hostile primes make judgments less hostile. Opposite theory-based correction would also occur if a person has different theories about how any given context influences judgments for different targets
(e.g., a person might believe that winning the lottery would make everyday activities seem more enjoyable than would otherwise be the case, but believe that winning the lottery would make one's current job salary seem unacceptably small). Although some models of assimilation and contrast predict that different effects on target judgment might occur for the same context (e.g., for some targets, the context might be "included" in the representation of the target which leads to assimilation, but the same context might be "excluded" from the representation of another target, which leads to contrast; Schwarz & Bless, 1992a), only a theory of correction based on judges' naive theories of bias predicts different corrections (i.e., away from and closer to the context) when the same context is believed by judges to have different biasing effects. Opposite corrections for the effects of the same context on different targets would present a unique problem for models of correction based on partialling or subtraction processes. This is because contextually activated reactions (that are subtracted in order to correct assessments of targets within those models) are the same for each set of targets and therefore, unless the contextual reactions are rather mixed, subtraction
of overlap will tend to move assessments of each target in the same rather than opposite directions.

In order to examine the viability of the critical opposite-correction outcome derived from the FCM, we first identified contexts and judgments for which people possessed opposite theories of how the same context influenced judgments about different targets (Wegener & Petty, 1995b, Study 1). Opposite theories for these contexts and judgments included a belief that perceptions of how desirable it would be to spend two weeks in very desirable vacation spots (e.g., Hawaii, Paris) would make staying in Midwestern cities (i.e., Indianapolis and Kansas City) seem less desirable, but would make jobs in the vacation locations seem more desirable. Respondents also believed that making ratings of attractive actresses (e.g., Michelle Pfeiffer, Kim Basinger) would make average people (i.e., Hillary Clinton and Tipper Gore) seem less attractive, but would make products endorsed by the actresses seem more desirable.

Meta-analyses of three initial data collections—each using a 2 (Theory of Uncorrected Effect: assimilation vs. contrast) × 2 (Correction Instruction: none vs. correction) design—suggested that opposite corrections consistent with a theory-based approach were occurring (Wegener & Petty, 1995b, Study 2). That is, when people believed that a given context would have an uncorrected effect of making targets seem less like the context than usual, corrections led target ratings to be closer to ratings of the context. When people believed that the same context would have the opposite uncorrected effect on (other) targets, however, corrections moved target ratings in the opposite direction.

These opposite corrections consistent with a theory-based view were also found in a single study in which participants were first asked for their context-independent perceptions of the targets (in order to control for baseline differences in perceptions of the targets; Wegener & Petty, 1995b, Study 3). That is, early in the research session, participants were asked to rate the experimental targets with no context items present. After completing some filler materials, participants encountered the context and target items from the vacation-location stimuli described previously and then encountered the context and target stimuli from the attractiveness context described previously. Participants either encountered these stimuli along with a correction instruction for each set of targets or with no correction instruction, and the target stimuli within both contexts were either the targets that people believed would seem less desirable within that context (i.e., the midwestern cities and average-looking people) or the targets that people believed would seem more
desirable within the context (i.e., jobs in the vacation locations and products endorsed by the actresses).

The ratings of the targets presented outside the judgment context were subtracted from the ratings of the targets inside the judgment context to form the primary dependent measure (i.e., the shift in ratings induced by the context and by correction processes). The same result occurred for each of the contexts participants encountered—the expected Theory X Correction interaction, $p < .0014$ (see Figure 4). For participants who received targets believed to seem too negative within that context (e.g., those rating average-looking people after rating very attractive actresses), the difference between ratings of context-imbedded targets and context-independent targets was more positive when a correction instruction was given than when no correction instruction was given. However, for participants who received targets believed to seem too positive within that context (e.g., people rating products endorsed by attractive actresses after rating

![Figure 4](image)

Fig. 4. Shift in target ratings as a function of theory of bias and correction instruction. (Adapted from Wegener & Petty, 1995b; Study 3.)
the actresses), the difference between ratings of context-embedded targets and context-independent targets was less positive when a correction instruction was given than when no correction instruction was given.

These data demonstrated a number of important results. For the first time, evidence that corrections can result in "overcorrection" leading to "corrected assimilation" of the targets was observed. That is, when people believed that a positive context would make targets seem too negative, correction instructions led to shifts in target ratings that became even more positive than ratings of the targets outside any context (i.e., the shift score under correction conditions was significantly higher than zero or "accurate correction": $p < .001$). The results of this study also nicely illustrated that corrections are made for the expected or perceived bias rather than for the actual bias present in the situation. Specifically, although participants significantly corrected in directions consistent with the shared theories of bias identified in a theory-identification study, they did so in a situation where the actual bias created by the contexts was negligible (perhaps because the number of contextual stimuli was smaller than in previous studies, see Wegener & Petty, 1995b).

Although corrections in the above studies presumably represent corrections according to the opposite theories of bias associated with the context\target pairs (as indexed in the theory-identification studies), it is possible that the differing corrections were due to the differing targets rather than the differing theories per se. Because target and theory both changed across conditions in those studies, one could gain further evidence for the necessity of differing theories of bias and for the flexibility of theory-based correction by demonstrating opposite corrections for the effects of different contexts on the same target(s). That is, if different contexts can be identified that are associated with different theories of bias for the same target, and if opposite corrections are observed consistent with those theories, then differential corrections cannot be the result of different targets rather than different theories. We have found evidence of opposite corrections of perceptions of the same target using two different sets of stimuli that differ markedly from the studies reported above.

In one study (Wegener, Petty, & Dunn, in press; Study 1), research participants rated the size of either large (e.g., cow, lion) or extremely large animals (e.g., whale, elephant) and then rated ambiguous animals (i.e., lemphor and monkey; see Herr et al., 1983). Ratings of the target animals were made either immediately following the context ratings or after a correction instruction (as in Petty & Wegener, 1993, see prior description). For the context expected by pretest subjects to make targets seem larger (that is, the moderately large context) the correction instruction tended to
make ratings of the target animals smaller. For the context expected by pretest subjects to make targets seem smaller (that is, the extremely large context), however, the correction instruction tended to make ratings of the target animals larger (Context X Correction Instruction \( p < .07 \)).

In a second study (Wegener, Petty, & Dunn, in press; Study 2), participants first rated either three extremely violent people (i.e., Adolf Hitler, Josef Stalin, and Saddam Hussein—a context expected by participants in a theory-identification study to make perceptions of targets less violent than usual) or three extremely nonviolent people (i.e., the Pope, Jesus Christ, and Gandhi—a context expected by participants in a theory-identification study to make perceptions of targets more violent than usual; cf. Herr, 1986). Then, research participants either immediately rated two target people (i.e., Arnold Schwarzenegger and George Foreman) on the same scale (No Correction condition), or were first asked not to let their perceptions of these target individuals be influenced by perceptions of the people they had just rated (Correction Instruction condition).

Results showed only the expected Theory \( \times \) Correction interaction, \( p < .0009 \) (see Figure 5). For participants who rated the targets in the extremely violent context (and expected that perceptions of targets would be biased toward less violence than usual), target ratings were more violent when a correction instruction was given than when no correction instruction was given. However, for participants who rated the targets in the extremely nonviolent context (and expected that perceptions of targets would be biased toward more violence than usual), target ratings were less violent when a correction instruction was given than when no correction instruction was given. In a control condition, people completed the target ratings before making any ratings of context items. These control responses (\( M = 5.57 \)) indicated that actual bias was only present under no-correction conditions when targets were rated after the extremely violent people. When targets were rated after the extremely nonviolent people, ratings did not differ from context-independent ratings. Thus, corrections away from initial perceptions of targets following the nonviolent context represent another case in which participants corrected according to their theory of bias even though no demonstrable bias was operating. In addition, corrections in the nonviolent context condition led target ratings to be somewhat lower than context-independent ratings (\( p < .06 \)). Thus, the results of this study cannot be accounted for easily by response language effects (see earlier discussion) because the response language perspective cannot account for a correction that occurs when no observable bias is present under “no-correction” conditions, and the corrected assessments of targets differ from the no-context ratings of the target.
Recently, opposite theory-based corrections of ratings of the same target were also obtained by Martin (1996). In Martin's study, participants rated context people who had been pretested to be either attractive or unattractive and then rated moderately attractive targets either with or without warning them that the context could bias their perceptions of the targets. Pretest subjects verified that people perceived the attractiveness context as having an uncorrected effect of contrast (i.e., an attractive context would make the targets seem less attractive and the unattractive context would make the targets seem attractive). As in our studies guided by the FCM, Martin (1996) found contrast when there was no warning of bias, but found assimilation when participants were warned of the potential bias.

F. DO PEOPLES' IDEOGRAPHIC THEORIES OF BIAS PREDICT CORRECTIONS?

In the initial tests of the FCM, there was no evidence that the participants in the correction studies held the same theories of bias as participants in
the theory-identification studies. Therefore, although such an assumption seems reasonable, we had not shown that an individual's theory of how a context affects his or her own perceptions of the target drives corrections for the context. An interesting implication of considering an individual's theory of the biasing effect of a context is that peoples' theories likely vary along both direction and magnitude dimensions. Thus, in assessing individuals' theories of bias, one might find evidence for differences in magnitude of correction as well as direction. That is, consistent with the FCM, the greater the biasing effect an individual believes some context has had, the greater the magnitude of correction that should be undertaken.

In one study examining this idea (Wegener & Petty, 1995b, Study 4), participants' unique theories of bias were measured prior to their making judgments under conditions encouraging correction of target ratings for perceived context-induced bias. With this procedure, we could examine whether participants corrected for the magnitude and direction of their own theories of how the context influences target perceptions. If individuals' theories of bias drive corrections, then regressing participants' corrected target judgments on their theories of bias should yield a significant negative relationship. That is, the more negative (or less positive) a person's theory of how the context might bias his or her perceptions of the target, the more positive (or less negative) the person's shift in target ratings should be when attempting to correct for the perceived bias. Thus, negative theories of bias should lead to more positive corrections of target ratings than should positive theories of bias (opposite directions of correction as in the studies described above). Furthermore, within valence of theory (e.g., within negative theories) more extreme theories should lead to larger corrections (i.e., adjustment commensurate with the perceived magnitude of bias).

In this study, the primary dependent measure was the shift in target ratings from baseline (context-free) ratings to those in the experimental setting. Theories of bias significantly predicted shifts in target ratings ($B = -.592$, $p < .0009$), indicating that as theories of bias became more negative, shifts in target ratings became more positive. Within the overall effect, there was evidence of correction associated with both direction and magnitude of the perceived bias. That is, a dichotomous variable denoting direction of perceived bias accounted for significant variance in the positivity of shifts in target ratings ($B = -.752$, $p < .0085$). In addition, participants corrected to a greater extent as their theories of bias became more extreme, controlling for the direction of bias (see Wegener & Petty, 1995b, p. 47 for descriptions of the various magnitude analyses).
G. SUMMARY

Our initial set of studies on the FCM provided evidence for a variety of flexibilities in corrections of social judgments. Our theory-identification studies showed that people can believe that either assimilation or contrast can be the uncorrected impact of contexts, and even showed that people can possess or generate opposite theories of bias for the same contextual factor influencing different targets of judgment or for different contexts influencing the same targets. Our correction studies provided the first empirical evidence of corrections based on judges' naive theories of bias. Across a variety of contexts and types of judgments (e.g., desirability of locations or weather, attractiveness of people or products, violence of people, size of animals), shared theories of bias found in the theory-identification studies predicted corrections when either explicit or more subtle prompts encouraged corrections of target ratings. Some of the studies provided the first evidence for assimilation effects resulting from theory-based corrections (i.e., "overcorrections") for perceived contrast.

In some of our studies, people had opposite theories of bias associated with different targets of judgment (within the same context). In other studies, opposite theories of bias covaried with the judgment context, and in still other studies, opposite theories occurred across individuals when target and context were held constant. In each case, opposite theories of bias predicted opposite corrections. Furthermore, the magnitude as well as the direction of participants' reported theories of bias predicted later shifts in target ratings under conditions that encouraged corrections. None of these predictions or results were derivable from the partialling views of bias correction (i.e., the set¬reset and inclusion¬exclusion models; see Petty & Wegener, 1993; Wegener & Petty, 1995b, for additional discussion of the partialling models).

IV. Corrections across the Many Domains of Social Judgment

One important asset of the flexible correction perspective is that its implications extend beyond the typical paradigm used in our initial studies (e.g., see discussion in Petty & Wegener, 1993). That is, when any perceived biasing factor is noted and the "unbiased" qualities of a target are sought, the correction process driven by theories of bias could proceed in a similar manner across many different domains of social psychological inquiry. Such correction situations are potentially present in many domains of inquiry—including persuasion, attribution, impression formation, stereotyping, and
more. In the following sections, we review some recent work across a variety of areas that can be understood using the FCM. In many of these areas, the FCM makes unique predictions and focuses the research questions in different ways than the dominant models in those areas.

A. PERSUASION

Although much has been learned about the processes of persuasion over the last 50 years (see Eagly & Chaiken, 1993; Petty & Wegener, in press), little attention has been given to the potential role of correction processes in persuasion settings. Nevertheless, as suggested at the beginning of this chapter, there might be many persuasion settings in which corrections for various perceived biases could occur. People might wish to correct for perceived biases associated with factors irrelevant to the presentation of the persuasive appeal (e.g., mood of the message recipient, Petty et al., 1993; Wegener & Petty, 1996b; see later section on mood and correction) or for factors more closely associated with the presentation of the persuasive communication (e.g., characteristics of the source who presents the message, Chaiken, 1987; Petty & Cacioppo, 1986).

To date, no explicit treatments of corrections for perceived biases have appeared in the persuasion literature (but see Petty, Priester, & Wegener, 1994). In fact, some discussions have included predictions that appear to run counter to the concept of bias corrections (at least in some instances). Consider, for example, the influence of source characteristics on persuasion. Although many factors (such as personal relevance/importance of the message topic) have been shown to moderate the impact of source characteristics, relatively little is known about the particular processes by which these factors have their effects. One hypothesis (within the Heuristic-Systematic model of persuasion [HSM]; Chaiken et al., 1989) is that source characteristics (such as how likable the message source is) can influence attitudes by invoking a relevant stored judgmental heuristic (e.g., “I agree with people I like”). Eagly and Chaiken (1993) further noted that such heuristics influence persuasion only to the extent that they are available and accessible in memory and that “factors that affect the accessibility of heuristics should exert a corresponding effect on the judgmental impact of heuristic cues” (p. 330). That is, heuristics should have a greater impact on attitudes to the extent that the heuristics come easily to mind when the message recipient encounters a persuasive appeal. Also, this model contends that as long as the invoked heuristic does not contradict the message arguments presented, it should add to the impact of the arguments when motivation to process is high (Maheswaran & Chaiken, 1991). This position has generally received
empirical support and seems especially useful in contexts where multiple possible heuristics might be hypothesized to exist (e.g., heuristics concerning multiple source factors such as likability, attractiveness, and expertise, or message factors such as length or number of arguments).

However, if one considers the potential role of correction processes in persuasion, an alternative possibility comes to mind. That is, when heuristics are made salient, people might perceive at least some heuristics (and the peripheral cues on which they are based, such as source attractiveness) as introducing unwanted biases into the persuasion setting, and thus people might attempt to remove the influence of the heuristic cue assuming people have the requisite motivation and ability to engage in corrective action (see Petty, 1994; Wegener, 1994). Of course, communication norms and other factors in the persuasion setting might also make some decision rules seem less than legitimate even though the decision rule itself might have formed because of some perceived validity of the rule (at least in some contexts).

We have conducted a number of recent studies that have provided support for the utility of the FCM in persuasion settings. In one study, for example, Petty, Wegener, and White (in press, Study 1) examined the extent to which people would correct for the likability of the message source when the potentially biasing aspect of this source information was made salient. Students received a proposal from a likable or a dislikable source that contained either strong or weak arguments in favor of changing a university regulation. Following receipt of the message, participants either immediately completed attitude measures, or were asked not to let any biases toward the message source influence their judgments of the proposal. When no correction instruction was given, attitudes toward the proposal were more favorable when the message was presented by a likable rather than dislikable source. However, when a correction instruction was given, attitudes were actually more favorable when the message was presented by a dislikable rather than likable source, suggesting an overcorrection for the perceived biasing influence. It is important to note that the correction instruction had no impact on the size of the argument quality effect on respondents' attitudes, indicating that the observed correction was not due to changing the amount of elaboration of the message content (Petty et al., 1976). The fact that an argument quality effect was present regardless of whether a correction occurred or not suggests that effort scrutinizing

\footnote{This view is compatible with the Elaboration Likelihood Model of persuasion (ELM; Petty & Cacioppo, 1986), which contends that when motivation and ability to process are high, people scrutinize all available information (including any invoked heuristics) for validity, and thus cues that survive scrutiny and can act as arguments would add to the impact of any other stated arguments, but cues that were found lacking (e.g., inappropriately biasing) would not (see Petty, 1994; Petty & Wegener, in press).}
message content is conceptually distinct from effort aimed at theory-based removal of bias.

Although we generally regard theory-based corrections as requiring greater cognitive effort than lack of correction (at least until such corrections become well practiced and routinized, cf. Smith, 1989), this is not to say that uncorrected outcomes are necessarily based on low levels of cognitive effort. That is, effort given to theory-based corrections is conceptually distinct from effort given to the task in general. Recall that the FCM suggests that theory-based corrections occur only if people identify a factor as potentially biasing and are then motivated and able to remove the perceived influence of that factor. Cognitive effort can often be given to a task when needs for bias correction are not salient and no operating bias is identified. Consider, for example, effort given to the scrutiny of information in a persuasive appeal. If a high level of scrutiny is given to information central to the merits of a position or object, and relatively little attention or weight is given to information irrelevant to the central merits (e.g., whether the source of the information is attractive or not), then it might be relatively unlikely that this irrelevant piece of information would be identified spontaneously as potentially biasing. If so, theory-based corrections would also be unlikely. If, however, the potentially biasing nature of that irrelevant information is made salient to message recipients (as in the Petty et al., in press, experiment), then the likelihood of theory-based corrections would be increased. Importantly, the bias could be made salient regardless of the amount of effort that went into scrutiny of the message content.

Such a perspective suggests that there might be multiple routes to diminishing the impact of peripheral cues. That is, the impact of peripheral cues could be negligible because high levels of scrutiny of attitude-relevant information overwhelm the potential impact of the cue (Chaiken et al., 1989; Petty & Wegener, in press), or because of explicit corrective effort to remove a perceived bias associated with the cue (Petty, 1994; Petty & Wegener, in press; Wegener, 1994). Of course, the theory-based correction view also raises the interesting possibility that effects of cues could be reversed under conditions that encourage corrections.⁷ Studies guided by

⁷ It is also important to note that multiple routes to diminishing the impact of biasing factors are also likely to exist in other research areas. For example, corrections might be responsible for lack of impact of false information (e.g., Golding et al., 1990) or for conscious overriding of stereotypic responses (e.g., Fazio et al., 1995). Of course, it is also possible, however, that effortful processes such as searching for and scrutinizing large amounts of diagnostic, individuating information overrides the biases (see Fiske & Neuberg, 1990) without necessarily involving corrections of perceived biases. Thus, one goal for future research will be to investigate which settings involve corrections for perceived bias and which do not. The possible correction-based outcome of reversing the effect of biasing factors is likely to be helpful in such efforts.
the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1979, 1986) and the HSM (Chaiken et al., 1989) have generally found decreasing impact of peripheral cues as elaboration of attitude-relevant information increases. The fact that studies guided by these frameworks have never reported reversals of the usual impact of cues suggests that high levels of scrutiny (such as that induced by high personal relevance; Petty & Cacioppo, 1979), might not typically induce efforts at correction for perceived bias associated with the cues (although corrections could have been responsible for attenuation of cue effects if the message recipients’ theories of bias associated with the cues were reasonably accurate).

Recently, Petty et al. (in press, Study 2) provided the first examination of correction processes across different levels of initial message scrutiny. In this experiment, research participants received a strong message on a topic that was made to be either high or low in personal relevance to message recipients. The message was presented by either a likable or dislikable source and was either followed by an instruction not to be influenced by any biasing characteristics of the source or not. In the no-correction conditions, a pattern consistent with past studies guided by the ELM and HSM was obtained—source likableness had an impact under low but not high relevance conditions (e.g., Chaiken, 1980; Petty et al., 1983). When people who had engaged in little processing of the message (i.e., low relevance conditions) received a correction instruction, however, the source likableness effect disappeared. This was likely due to explicit correction processes, given that a similar correction instruction did not affect the extent of message processing in the first experiment by Petty et al. (in press).

Was the lack of cue impact in the high elaboration no-correction conditions observed in this and prior studies due to explicit corrections for likability, or due to scrutiny of the central merits of the arguments overwhelming the cue? If high levels of scrutiny explicitly involve attempts to remove the impact of cues, then the correction instruction (not to be influenced by source characteristics) in the high elaboration conditions should have little further effect. If, however, message recipients did not consider the biasing nature of the cue while effortfully processing the central merits of the message arguments, then a correction instruction might have just as much impact as in the low-processing-correction condition. In fact, when people who had initially processed the message at high levels received a correction instruction, the impact of the cue was reversed such that the dislikable source was more persuasive than the likable source. Thus, it appears that enhanced processing motivated by high personal relevance does not imply correction, and that corrections can occur regardless of the effort initially put into a judgment task. The fact that corrections occurred even when there was no impact of the cue before correction (i.e., in the
high-processing conditions) again suggests that corrections are aimed at removing the perceived effect of the biasing factor rather than the actual bias that occurs.

Of course, theory-based corrections can also occur for perceived biases that do not stem from source characteristics. For example, if people perceive that a factor in the persuasion context is affecting their reactions to the persuasive appeal, such as environmental noise making one's judgments more negative (Nisbett & Wilson, 1977), they might attempt to correct assessments of the target based on their perceptions of the biasing effect of the environmental factor. Note that in this case, the environmental noise is unlikely to be incorporated into a decision rule regarding validity of the message. That is, there is no decision rule that people should believe, or disbelieve, messages that are accompanied by white noise. To examine this possibility, in one study, Wegener, Edwards, Petty, and Weary (1996, raw data) played an audiotape of a persuasive message on which static made reception of the message difficult. After the message, all recipients received a correction instruction and were asked to provide their perceptions of the speaker and of how persuasive and convincing the message had been. The primary finding was that theories of bias associated with how white noise would affect perceptions of speakers and messages (which had been assessed during a “separate study” earlier in the session) predicted respondents’ ratings of the speaker and message. To the extent that people believed that white noise would make people react negatively (rather than positively) to a message or a speaker, ratings of the message and of the speaker were more favorable (rather than unfavorable).

Our persuasion correction work suggests that corrections occur across a variety of settings that differ in many ways from the survey-like social judgment paradigm used in many of the initial tests of the FCM. Moreover, consideration of bias correction in the persuasion setting generates research questions that address fundamental questions in the persuasion literature and take that literature in some new directions. Application and development of the FCM within other research domains also seems capable of moving those research endeavors forward.

B. ATTRIBUTION

Correction-like processes have long been a part of observation and theorizing in the literature on causal attribution. Specifically, across a variety of judgments, augmenting or discounting of initial causal attributions have been postulated to take place when additional causal factors become salient. Thus, when people consider a particular explanation for another person’s
behavior, their initial perceptions of the role of a given causal factor are presumably changed in order to take into account the role of any additional salient causal factors (Kelley, 1972, 1973). Adjustments of attributions as a result of salient contextual factors have been shown for judgments ranging from the causes of a person's behavior (e.g., Himmelfarb & Anderson, 1975) to assessments of a person's own internal reaction to various events (e.g., Fazio, Zanna, & Cooper, 1977; Zillmann, 1983). Although some perspectives on attribution have dealt with nearly the whole of attribution as an effortful process (e.g., Kelley, 1967), some more recent accounts assign greater cognitive effort to processes that either have been labeled explicitly as attributional corrections (e.g., Gilbert et al., 1988) or could be considered as such (e.g., Quattrone, 1982; Trope, 1986).

For instance, Gilbert and his colleagues discuss the attribution process as consisting of three steps: (a) categorizing the behavior (i.e., what is the actor doing?) (b) characterizing the actor (i.e., what trait does the action imply?), and (c) correcting the dispositional attribution (i.e., what situational constraints might have caused the action?). Importantly, Gilbert et al. (1988) noted that "attributions are a product of dispositional inferences that are followed by situational adjustments” and “correction is a species of reasoning (a higher order process), whereas characterization is a species of perception. . . . If this is so, then the correspondence bias can be seen as the failure to apply an inferential correction to the initial dispositional perceptions that perceivers cannot help but have,” (p. 738; see also Gilbert & Osborne, 1989; Osborne & Gilbert, 1992). Thus, Gilbert and his colleagues assumed that dispositional inferences were the default (uncorrected) outcome of categorizing the behavior and characterizing the actor. The impact of situational information occurred as a result of a relatively effortful correction of an initial dispositional inference.

Similarly, Trope (1986; Trope, Cohen, & Maoz, 1988) discussed attribution as consisting of two steps: identification and inference. Within Trope’s view, identification of the behavior as fitting a disposition-relevant category takes place in a relatively spontaneous manner. That is, some behavior is quickly (almost effortlessly) categorized in terms of which trait it implies. After the trait implications of the behavior have been identified, however, situational information is used to make inferences of the true disposition of the actor. If the situation would generally tend to inhibit the identified behavior, the diagnosticity of the identified behavior is augmented (i.e., the person’s true disposition is viewed as even more like the behavior-implied trait), but if the situation would encourage the identified behavior, the diagnosticity of the identified behavior is reduced (i.e., the person’s true disposition is viewed as less like the behavior-implied trait). This view is also similar to that of Quattrone (1982), who conceived of attribution as
an anchoring and adjustment process (in which initial perceptions of the actor's behavior are often anchored on the trait implications of the behavior, and then are insufficiently adjusted; for additional discussion, see Wegener, 1994).

Interestingly, just as pre-FCM tests of context correction effects used contexts for which assimilation was the natural effect (e.g., Martin et al., 1990), it is possible that initial tests of attributional correction models tended to use scenarios for which trait inferences were more natural than situational considerations. As a result, theories of attribution have tended to assume that dispositional inferences are the default, and that corrections lead away from dispositional inferences to take situational constraints into account. This does not preclude, however, that stimuli exist for which a situational attribution would come to mind before any consideration of trait qualities (see Anderson, 1983, for a discussion of differing situations activating differing causal alternatives). Applying the notion of flexible corrections to attribution leads to the differing hypothesis that either situational or dispositional attributions can be the default (uncorrected) outcomes, and either situational or dispositional attributions can be the result of effortful corrective processes (see Petty & Wegener, 1993; Wegener & Petty, 1995b).

Such possibilities have important implications for interpreting recent attribution studies. For example, Morris and Peng (1994) found that Chinese attributors reported more influence on social behavior by external factors and less influence by internal factors when compared with American attributors. Also, English-language newspapers provided more dispositional explanations and less situational explanations than Chinese-language newspapers when reporting the same crimes (for conceptually similar results using Hindu respondents, see Miller, 1984). Within the view of attribution espoused by Gilbert et al. (1988), this could be because of effortful correction generally taking place for Chinese but not American attributors. According to flexible correction notions, however, either dispositional or situational attribution could be the uncorrected outcome, and either dispositional or situational attribution could be the result of effortful correction processes. Thus, in addition to the view of Chinese attributors as chronically effortful correctors—generated from the Gilbert et al. (1988) model, the flexible correction view also suggests that something common to Chinese attributors (e.g., the Chinese culture) might have made situational attributions the default (no-correction) outcome that can subsequently be corrected in light of perceived dispositional influences.

In addition, it is likely that situational factors within the same culture could make situational attributions the default (uncorrected) outcome on some occasions. That is, even with American attributors, some settings
might encourage situational attributions for behavior, with information about the person only having an impact when attributors are able to effortfully consider the implications of the person's dispositions. For example, Petty and Wegener (1993) discussed the following situation. Consider a target person who is wearing fatigues and a helmet as he runs across a battlefield firing a machine gun. Although it is possible that social perceivers would first infer trait hostility before taking into account the situational constraints of being in a war, it seems more likely that situational explanations might be the first causal alternatives to come to mind, and that these alternatives could then be adjusted in light of subsequent dispositional factors that are considered.

That is, if a social perceiver realizes that his or her initial perceptions are being biased in the direction of the salient violence of the situation, these perceptions of bias might lead people to correct by attending to dispositional qualities of the person in that situation (Petty & Wegener, 1993; Wegener & Petty, 1995b). Thus, although the form of the theories of bias might differ somewhat from the context-effect to the attributional setting, the general tenets of the FCM also seem applicable to attribution theorizing.

In fact, support has recently been obtained for the FCM predictions of variation in uncorrected attributions and variation in direction of attribution corrections. For example, Krull and Dill (1996) found that focusing judges on the situation surrounding behavior made judgments of the situation faster than judgments of the person's disposition, but focusing judges on the person made judgments of disposition faster than judgments of the situation. This suggests that focusing people on the situation made situational attributions the default, but focusing people on the person made dispositional attributions the default. As noted by Petty and Wegener (1993), we believe that some judgment targets and settings might also make situational attributions the default, even if people have not been explicitly told to focus on the situation (for recent evidence, see Lupfer, Clark, Church, De Paola, & McDonald, 1995; Read & Yang, 1996).

Evidence for corrections away from initial situational attributions were found by Wegener and Petty (1995a, also reported in Wegener, 1994). This experiment was designed as the mirror image of the evidence for attribution corrections presented by Gilbert et al. (1988). That is, in this experiment, research participants heard a verbal description of the violent actions taken by a soldier on a battlefield. Before receiving the description, participants were either told that the person had chosen to disobey an order to cease fire or had chosen to obey an order to attack. During presentation of the battle description, participants were either asked only to assess the qualities of the situation in which the target had been placed or were asked to assess
the qualities of the situation and try to remember as many three-syllable words from the passage as they could. Asking participants to remember the three-syllable words or not varied the cognitive load of the participants.

Results showed that participants who were distracted by trying to remember words from the passage rated the violence of the situation in which the soldier had been placed as quite violent regardless of whether the person had chosen to obey or disobey the respective order, whereas participants who were not distracted rated the violence of the situation as higher when the person chose to obey the order to attack than when the person chose to disobey the order to cease fire. Thus, it appears that at least some situations exist in which assessments of situational qualities are adjusted in light of information about personal choices (but only when sufficient cognitive resources are available).

Krull and Erickson (1995) also presented evidence of flexible attributional corrections. In this research, participants viewed a silent videotape of anxious behavior in an interview with the goal of assessing the amount of anxiety provoked by the interview topic. Participants were also given information that the person’s personality was either calm or anxious. During the videotape, participants heard a series of tones. In the high cognitive load conditions, participants were asked to monitor and record whether each tone was high, medium, or low in pitch, whereas participants in the low load conditions were told to ignore the tones. Consistent with the FCM predictions (Petty & Wegener, 1993), Krull and Erickson found that information about the actor’s personality had a greater impact on judgments of the anxiety provoked by the situation when participants were under low rather than high load. Krull (1993) provided similar evidence, but without providing information about the personality of targets. That is, adjustments away from initial situational attributions occurred when people were not distracted but not when people were distracted. In addition, when judges were not distracted, targets were viewed in more dispositional terms when people were given a goal of assessing the situation than when they were given a goal of assessing the target’s disposition. Given that no information had been presented on which to base corrections, the observed adjustments in judgment might have been guided by theories of how the initial judgments were biased by the salient situational or dispositional explanation for the observed behavior. Recently, Gilbert and Malone (1995) have also endorsed a view of more flexible attributional corrections in which either dispositional or situational attributions can be the default or the result of attributional corrections (see also Uleman, Newman, & Moskowitz, 1996).

Additional work in the area of attribution might also benefit from application of the FCM. For example, when people find out that statements presented in the attitude attribution paradigm were elicited under conditions
of low choice (i.e., that the target person was assigned which side of the issue to support; e.g., Jones & Harris, 1967), they might engage in corrections that are directed by theories of how the statements might influence their perceptions of the target's true attitude. When statements are found to be assigned rather than chosen, this might undermine the legitimate use of the statements as a basis for inferring the person's attitude. Unfortunately, people might not realize the extent to which the statements have influenced their perceptions of the person, so corrections for the information might often be incomplete. The role of theories of bias in attitude attribution should receive attention in future research.

C. IMPRESSION FORMATION

As we noted earlier, although partialling models of bias correction have been dominant within the impression formation literature in recent years (e.g., Martin, 1986), a more flexible theory-based model provides some interesting and new perspectives on this well-developed literature. One of the primary differences between the FCM and prior partialling models regards the stringency of the bias-related assumptions in the various models. Perhaps because the partialling models were developed specifically out of the priming/construct accessibility literature (e.g., see Martin, 1986), the partialling models made some rather strict assumptions regarding the nature of bias and the nature of the processes that must occur for correction to take place. For example, recall that bias was viewed as the default assimilation resulting from an overlap between reactions to the context (i.e., the priming stimuli) and reactions to the judgment target, and correction was the partialling (subtraction) of overlapping reactions (which was assumed to take greater effort than use of the overlapping reactions; see Figure 1). Because the partialling was of overlapping reactions, the assumption was that overlap (or at least perceived overlap) must occur before correction processes would be relevant. Given these considerations, in the partialling models, "default" judgments were those that were uncorrected and relatively low-effort.

Our treatment of these issues in the FCM has differed from the partialling models in several respects. Perhaps most importantly, focus on corrections guided by theories of bias removes any assumptions about the necessary process(es) that are responsible for the bias that is corrected. That is, within the FCM, no assumptions are necessary regarding the nature (or even direction) of the initial biasing effect.8 Regardless of what caused the initial

---

8 Consistent with the Flexible Correction view (e.g., Petty & Wegener, 1993), theorists in a number of areas have recently taken positions that various outcomes can be "default" (uncorrected) outcomes of judgment processes (i.e., that there is often no universal default judgmental outcome; e.g., Gilbert & Malone, 1995; Higgins, in press; Uleman et al., 1996; see also Krull, 1993).
perceived bias (and people might often be unaware of what caused it—or people might perceive it when it doesn’t really exist), corrections are guided by naive theories of the bias at work. Thus, in the FCM, “default” judgments are “uncorrected,” but they are not necessarily low in effort. Our first empirical demonstration of this was the Petty et al. (in press) persuasion experiment in which corrections occurred equally after both high and low initial elaboration of a persuasive message.

Recently, Martin (1996) provided evidence that is also consistent with this reasoning. Recall that in one experiment described previously, Martin replicated our correction for perceived contrast using judgments of attractive or unattractive actors as a context followed by judgments of average-looking target actors. In a second study, Martin manipulated the cognitive effort given to the same judgment task with no “warning of bias” present. That is, participants rated the attractive or unattractive context actors followed immediately by the average-looking target actors (as in the “uncorrected” cells of the earlier study). In this second study, however, Martin (1996) manipulated participants’ effort for the “uncorrected” effects by creating differences in personal responsibility across conditions. In “high effort” conditions, participants were told that they were personally responsible for the judgment, whereas “low effort” conditions consisted of telling participants that they were not personally responsible, but were simply part of a group making the judgment (and that a “top of the head” response was appropriate). In Martin’s (1996) initial study (described earlier—with no manipulations of effort), the “uncorrected” effect had been contrast (as in our studies). Martin found that decreasing the effort given to the task changed the direction of the effect from contrast to assimilation, but that the personal responsibility instruction (i.e., high effort conditions) resulted in contrast (as in the earlier studies). Although Martin interpreted these data as showing that a reset contrast effect (which requires effort) can then be overcorrected in the opposite direction when a “warning of bias” is present (shown in the Martin, 1996, study discussed earlier), we believe that the present study might instead show that there can be multiple “uncorrected” outcomes produced by various processes—any of which can then be modified by theory-based corrections. That is, if a reset correction occurred in the high-effort condition, it is unclear why an opposite overcorrection occurred when a “warning of bias” was introduced (in the study described earlier). It seems as (or more) likely that some kind of comparison process required effort (and brought about the initial contrast effect) but that the context was not identified as potentially biasing until the “warning of bias” was introduced. In some ways, the most interesting result of this

---

9 Although we do believe that a person can perceive an overcorrection as creating a bias that is then addressed in a second theory-based correction, it seems unlikely that the second correction would also be a significant overcorrection that would reinstate the same size of original bias that started the correction process.
study was the assimilative bias in the low-thought conditions. Although Martin did not include a correction instruction in that study, corrections in that setting would be interesting—especially if people still hold the “contrast” theory of bias that is reported for the general paradigm (when no personal responsibility component is included). If they do, a theory-based correction might actually exacerbate the assimilation bias observed with no correction instruction. If people hold a different theory for the situation in which low effort is introduced, however, then corrections might occur that diminish rather than exacerbate the assimilation bias.

When we began our work on theory-based correction (Petty & Wegener, 1993), one of the somewhat contentious issues raised by our perspective was that assimilation could be the result of effortful corrections (not just the “uncorrected” or “undercorrected” effects of overlap—as in the partialling models). Given that we, and now Martin (1996), have demonstrated “corrected assimilation” a number of times, perhaps the remaining contentious issue within our FCM framework is that contrast can be the uncorrected result and not just the “overcorrected” result of removing overlap. That is, Martin’s set–reset model might interpret his and our “uncorrected” contrast results as actually due to an initial reset correction that is then reverse overcorrected according to theories of bias. If so, it is not clear why the first “correction” is not theory-based but the second one is, or why people would overcorrect in the second correction. Although it seems more likely that a single theory-based correction followed some other judgment process during which no bias was initially identified, this double-correction alternative is possible. However, it does not seem reasonable to assume that all contrast effects require effortful removal of overlap. Recall Helson’s (1964) work described earlier. Perceiving warm water as cooler because of first placing one’s hand in hot water does not seem likely to be mediated by a subtraction (or partialling) of reactions to the hot context water. The warm water simply seems cool in comparison. Thus, in our view it seems unwise to make assumptions that do not allow for contrast to be a “default” (uncorrected) outcome.

Recent evidence relevant to the possibility of “default” contrast was presented by Winkielman and Schwarz (1996) who exposed research participants to subliminal presentations of either small or large animals (study 1) or small or large American states (study 2). When later judgments were of the size of mammals in general (study 1) or the size of states in general (study 2) assimilation occurred; but when later judgments were of the size of a specific animal or of a specific European country not primed by the subliminal exposure, contrast from the primed material was the result. It seems unlikely that contrast is the result of partialling perceived overlap between reactions to the target and reactions to a task of which participants
were unaware. Thus, this work seems to indicate that contrast can, in fact, be the “default” (uncorrected) result of exposure to a context (at least, it seems unlikely that a partialling type of correction could bring about the effect). It would be interesting to test whether participants might engage in theory-based corrections of such contrast effects if they were made aware of the nature of the subliminally presented priming stimuli.

In sum, the assumptions of the FCM differ from the partialing models in that the FCM holds that outcomes such as assimilation or contrast can be the result of corrections or lack of corrections rather than tying correction (and cognitive effort) to a particular outcome. As a whole, the FCM provides an interpretive framework for many contexts in which impression formation phenomena occur. In the following sections, we illustrate such areas using research on impression formation in courtroom settings and on the use of stereotypes as a specific source of bias in impression formation settings.

1. Courtroom Judgment

In one study of courtroom judgment, Thompson et al. (1981) presented mock jurors with videotapes of evidence in a trial. In one condition, no inadmissible evidence was presented, but across the other two conditions, the inadmissible evidence was either proconviction or proacquittal. Inadmissible evidence that was proacquittal decreased perceived guilt of defendants, but inadmissible evidence that was proconviction did not significantly influence perceptions of guilt. Thompson et al. (1981) speculated that people might want to correct for the influence of proconviction more than proacquittal inadmissible evidence, and also speculated that perceptions of bias might mediate those effects (although assessments of perceived influence that were collected after the evidence was provided and deliberations took place showed that “jurors” believed that the proconviction evidence had influenced them but the proacquittal evidence had not—the opposite pattern of the observed guilt effects). Although posttask assessments of perceived bias might be difficult for subjects to provide and to interpret, it might also be possible that the assessments of perceived influence in the Thompson et al. study were indexing perceptions of mock jurors that had driven corrective attempts. That is, if jurors perceived the inadmissible proconviction evidence as influencing them as the evidence was presented, they might have engaged in more corrective effort than when inadmissible proacquittal evidence was presented. This would create the judgment effects that were observed. What the Thompson et al. posttask ratings might point out, however, is that people are limited in their ability to accurately perceive
the extent to which their attempts at correction have been effective (when looking back at past judgments).

Further investigation of the role of theory-based correction in courtroom judgments can proceed in many of the same ways as it has for other social judgments. Thus, for example, one could identify different types of courtroom evidence for which people hold different theories of how the evidence would influence their perceptions of the defendant. People might even hold different theories about evidence for different kinds of crimes or concerning different classes of defendants. In addition, different people might believe that the same evidence has affected their views of the defendant in different ways. If theory-based correction takes place in such settings, one should find that the differing theories of bias predict different corrections when the evidence is discredited.

In our initial study of this type (Wegener & Petty, 1996a, raw data), we provided research participants with a brief description of a hypothetical case in which the defendant was charged with rape. In one version of the description, a witness stated that he had previously known the defendant as a fellow prisoner convicted for assault. In another version, the previous conviction was for tax evasion rather than assault. After assessing initial perceptions of likelihood of guilt based on this information, research participants received an additional description noting that another witness came forward and disputed the previous testimony as untrue (which led to a recanting of the previous testimony). The difference between perceptions of guilt before and after the recanting of the testimony was taken as an indication of the correction for the perceived effects of the testimony. Weeks before the study, participants had provided their perceptions of the likely effects of different kinds of testimony and provided baseline ratings of perceived likelihood of guilt for a case description identical to the description in the study (but without the prior-conviction testimony). Thus, we had assessments of the actual impact of the prior conviction testimony and of perceptions of likely bias associated with it. Although research participants believed that the prior conviction for assault would affect perceptions of guilt more than prior conviction for tax evasion, the effects of these two testimonies actually did not differ. Importantly, however, the corrections that occurred when the testimony was recanted did reflect the differences in perceived bias that were associated with the two types of testimony (i.e., people corrected for the perceived effects of an assault prior conviction, but did not correct for a tax evasion conviction—presumably because they did not think that the information had affected their assessments of guilt). Thus, this study provides another indication that perceptions of bias (separable from the experience of bias) drive corrections.
In the study just described, people corrected for the prior assault information because they believed this information had affected their judgments and it was appropriate to correct for it because this biasing information was proven false. In actual court cases, people are sometimes expected to correct for certain biasing information even if it is true as long as this prejudicial information is deemed inadmissible by the court. For example, information about a defendant's prior crimes would generally be deemed inadmissible as evidence. Yet, jurors might sometimes feel that using this information would help them reach a more accurate verdict (i.e., convicting the right person).

This type of situation is a fertile one for examining the idea that people sometimes acknowledge that they have a bias, or some information has produced a bias, but that the bias holder believes that this bias is relevant to producing an accurate, or at least an appropriate judgment. In such cases, people should be less likely to correct for the bias than when the bias is seen as clearly illegitimate. In one study examining this hypothesis, Petty, Wegener, and Fleming (1996) had college students read about a court case in which a person was accused of the crime of rape. There were three versions of the case materials. In the control condition, the students were provided with facts that pretesting indicated would lead them to be uncertain as to whether or not the accused had committed the crime. In two other conditions, students read the control information but also learned that the suspect had pled guilty to a prior rape or a prior burglary. Although people are generally biased against those who have committed crimes when they are making snap judgments, we hypothesized that when trying to be a good juror, they would correct for any bias due to the prior burglary but not the prior rape because the latter would be seen as more relevant to a proper verdict. In this study, after students had read the appropriate case materials, they gave two judgments. One judgment simply asked them to indicate what they really felt about the defendant outside the jury setting. On this measure, both the defendant who had committed the prior burglary and the defendant who committed the prior rape were seen as more guilty than the defendant in the control condition who had no previous criminal background. That is, committing any previous crime inflated perceptions of guilt, perhaps because students have a naive theory that some people are just a "criminal type." On a second measure, however, the students were asked to rate how they would vote if they were an actual juror in the case. This measure would presumably motivate people to remove any unwanted or inappropriate biases from their judgments. Consistent with the legitimacy hypothesis, on this measure, the burglary defendant was no different from the control defendant in rated guilt, but the
defendant who had committed a prior rape was seen as more guilty (see Figure 6). That is, when presumably attempting to be a good juror, participants corrected for what was potentially an illegitimate bias (i.e., a prior burglary conviction), but did not correct for a more relevant bias (i.e., a prior rape conviction). Thus, it might be the case that evidence that is ruled as inadmissible, but still relevant to making an accurate assessment of guilt (and thus seen as legitimate), might continue to be used by jurors even if they are instructed to disregard it (e.g., see Thompson et al., 1981; Kassin & Sommers, in press).

This legitimacy principle can also explain the results obtained by Golding et al. (1990). In this research, information about a target person had no impact on impressions of the person when the information was later labeled as incorrect (i.e., false), but had an impact when it was later labeled as confidential, and not to be used. It is possible that the research participants in this study believed that the implications of confidential information for
judgment were legitimate to use because they were valid and relevant to making accurate judgments of the target person even though they were asked not to use the information because of its confidentiality. In the case of false information, however, any implications of the information were not legitimately attributed to the target person, and because of this, judges were motivated to correct their assessments of the target away from the implications of the false information.  

2. Stereotyping/Prejudice

Perceived legitimacy of a bias is also likely to influence corrections for biases based on stereotypes or prejudice associated with social groups. For example, both people who score high and those who score low on measures of prejudice or racism might be equally aware of the bias contained in the stereotype of a social group (e.g., Devine, 1989), but high scorers might view the evaluative bias in the stereotype as more legitimate than low scorers. Because of this, high scoring perceivers might be unmotivated to correct their judgments away from the implications of stereotypic information even though they know that the information will make particular perceptions of the target more likely than if the information were not considered (i.e., that the information is “biasing”).

In a recent development in this area, Fazio and his colleagues demonstrated that individual differences in perceived legitimacy of prejudice-based biases moderates prejudice correction phenomena (Dunton & Fazio, 1996; Fazio, Jackson, Dunton, & Williams, 1995, Study 4). Fazio and his colleagues began by identifying people who varied in their motivation to control prejudice (Dunton & Fazio, 1996). These individuals engaged in the Fazio et al. (1995) priming procedure, which involves presentation of African-American and Caucasian faces as primes during a task in which participants rate target adjectives as good or bad. The procedure is designed to measure attitudes toward African-Americans in an unobtrusive manner. Modern Racism Scores (MRS; McConahay, 1986) reflected a significant interaction between participants’ motivation to control prejudice and their attitudes as identified by the priming.
procedure. Whereas the MRS responses were consistent with unobtru-
sively measured attitudes for people low in motivation to control prejudice
(i.e., people with more negative attitudes were classified as more racist),
MRS scores were inversely related to attitudes for people high in
motivation to control prejudice.

If effects of motivation to control prejudice were not corrective in nature,
one might have expected that people with equally high motivation to control
prejudice would act in similar ways toward African-American targets or in
response to questions about African-Americans. However, the fact that
people who experienced higher levels of negativity in the priming procedure
actually provided controlled responses that were more favorable than peo-
ple who experienced lower levels of negativity suggests that the phenome-
non is corrective in nature. Within the FCM, people who are high in desire
to control prejudice but experience negativity when they encounter African-
American faces might be more likely to perceive that ethnicity has a nega-
tively biasing effect (and thus are more likely to attempt to correct for
the perceived effects of ethnicity). Equally concerned people who only
experience favorable reactions toward African-Americans should be less
likely to engage in corrections. The paradoxical effect of these tendencies
is that people experiencing negative reactions but correcting for them might
end up acting in a less prejudiced way than people who only experience
favorable reactions.

In another relevant study, Lambert, Khan, Lickel, and Fricke (in press)
showed that the mood state of social perceivers can influence corrections
for stereotypes and/or prejudice associated with the target. They hypothe-
sized that sad mood would increase motivation to correct for perceived
bias relative to neutral mood, and that a happy mood would decrease this
motivation (cf. Schwarz, 1990). In this research, when perceivers were
prejudiced against a group to which the judgment target belonged, Lambert
et al. found that sad moods led to overcorrection (i.e., viewing the target
more favorably when she belonged to the disliked group than when she
did not), but happy mood led to greater use (lack of correction) of the
negative opinion of the group than a neutral mood. When perceivers liked
the group to which the judgment target belonged, however, no corrections
took place. Consistent with the FCM, Lambert et al. explained this by
noting that the negative attitude toward the group was likely viewed as an
inappropriate bias and was corrected according to a theory of bias associated
with that attitude. In contrast, positive attitudes toward the group were
assumed to be viewed as appropriate and thus were not corrected. The
hypothesized mood-based differences in motivation to correct (and the
mood-based differences in judgment outcome) are an interesting aspect of
this work that deserve additional research attention in the future.
Within the FCM, another way to make sense of these results would be to consider the potential effects of mood on perceptions of the biases in this setting. As in the Fazio et al. situation described earlier, people who are experiencing negativity (in this case, because of a negative mood) might be most likely to perceive the attitude toward the group as negatively biasing perceptions of the target person. Of course, this would happen to a lesser extent as feelings of negativity decrease from neutral to happy mood. Considered in a slightly different light, one might regard one's current mood as influencing the perceived applicability of a generally held theory about the biasing effects of the prejudiced attitude. That is, assuming that people regarded the prejudiced attitude as likely to inappropriately make them view the target person more negatively than would otherwise be the case (as Lambert et al. assumed), social judges might have viewed this theory as less likely to apply to this particular case to the extent that they were also experiencing positive (rather than negative) reactions as they encountered the target. Either of these FCM-generated possibilities could also create the differences in corrections for prejudice across mood states.\(^\text{11}\)

**D. MOOD AND OTHER SUBJECTIVE EXPERIENCES**

At a number of points in this chapter, we have discussed mood as a possible biasing agent that could be corrected through the theory-based processes of the FCM (see also Petty & Wegener, 1993; Wegener & Petty, 1995b). Much of the past work on mood and judgment has utilized noncor-

\(^{11}\) In addition to the correction differences across mood that might result from differences in perceptions of bias, management of mood states might also influence engagement in effortful corrective activities. Consistent with our hedonic contingency hypothesis (see Wegener & Petty, 1994; Wegener & Petty, 1996b; Wegener, Petty, & Smith, 1995), if a task in which corrections are studied is uplifting, happy people might especially engage in that task (including corrections). If, however, the task is depressing or would undermine the good mood, then happy people might be especially likely to avoid engaging in the task (and thus would also not engage in the corrections). People in sad or neutral moods might be less affected by the pleasantness or unpleasantness of the activity. Although no complete test of these possibilities exists, the task in the Lambert et al. studies might represent a portion of this picture.

If the tasks involved in the Lambert et al. studies were viewed by participants as capable of undermining mood, happy people might have engaged in the tasks less than neutral or sad people (e.g., by using a stored attitude toward the group to judge the target instead of scrutinizing the presented information). Importantly, as noted in our discussions of the hedonic contingency hypothesis (e.g., see Wegener & Petty, 1994; Wegener et al., 1995), the hedonic contingencies per se could also create differences in task engagement between neutral and sad moods. One important question for future research is when such differences emerge and when factors beyond the hedonic contingencies themselves eliminate differences between neutral and sad moods (as in Wegener & Petty, 1994).
rective models (e.g., the mood-as-information view, Schwarz & Clore, 1983). That is, effects of mood have been attributed to use of mood as an answer to a “how do I feel about it?” heuristic (Schwarz, 1990), and undermining the “informativeness” of mood has led to a lack of effect for mood (because judgments are no longer based on the feeling heuristic, but are rather based on judgment-relevant information that does not differ across conditions—typically because subjects have been randomly assigned across moods). However, if one takes a flexible correction view of situations in which people become aware of the true source of feelings (rather than attributing them to the target of judgment), it should also be possible to show differences in judgmental outcome, depending on the operating theories of bias for the given judgmental setting. For example, although the mood-as-information view does not provide a ready mechanism by which contrast away from moods might occur, overcorrection based on exaggerated theories of bias provides such a mechanism. Interestingly, a number of recent studies are consistent with such flexible correction predictions.

In one study, for instance, Berkowitz and Troccoli (1990) induced discomfort by asking some research participants to keep their arm extended and unsupported for a period of time. Other participants rested their arms on a table. During these activities, participants listened to an “autobiographical statement” by a target person. When attention was drawn to feelings (by asking people to rate their feelings before rating qualities of the target person), judgments of the target were actually more positive to the extent that feelings were more negative (a contrast effect away from mood). When attention was not drawn to feelings (because an innocuous distraction activity took the place of the feelings rating), judgments were more negative when feelings were more negative (a mood-congruent outcome). According to Berkowitz and Troccoli (1990), attention to feelings led to heightened cognitive activity and to greater attention to more peripheral cues in the situation that would suggest a desirable way of responding. However, according to the FCM, overcorrections based on naive theories of bias could be responsible for the mood-incongruent outcome when mood-based biases were made salient (i.e., in feelings-attention conditions).

Similarly, Ottati and Isbell (1996) induced a happy or sad mood before they presented people with information about a political candidate and told those people that they would be asked to rate that candidate at a later point in time. Results showed mood exerted an assimilative effect on judgments of the candidate when the judges were not politically expert, but contrast from the mood resulted when judges had high levels of political expertise. In order to interpret their results, Ottati and Isbell invoked flexible correction constructs, and argued that theory-based corrections occurred for political experts but not for nonexperts. That is, they distin-
guished their position from Martin's (1986) set-reset formulation by positing that adjustments in evaluations of the candidate were made "to compensate for the perceived biasing influence of mood" (p. 41). Of course, as noted previously, this is one of the key differences between the FCM and the previous set-reset view. Interestingly, although Ottati and Isbell did not measure theories of bias associated with mood, Petty and Wegener (1993) included a measure of perceived mood-related bias that showed people hold a belief in mood-based assimilation (for mood effects on perceptions of everyday activities). Thus, although there is little evidence in the Ottati and Isbell research that corrections for mood were actually driven by theories of bias, it is certainly plausible that this is the case.

A variety of subjective experiences other than mood could be identified by people as unduly biasing their judgments and requiring correction. For example, consider a recent study by MacDonald, Zanna, and Fong (1996) in which intentions to engage in unprotected sex were influenced by intoxication with alcohol. In order to rule out alcohol expectancy effects (i.e., belief that one was intoxicated "giving license" for otherwise troublesome behavior), a placebo condition was included in which people believed that they were consuming large quantities of alcohol even though actual alcohol content was minimal. Interestingly, within this placebo condition, the more intoxicated people thought they were, the less likely they were to intend to engage in unprotected sex. Why might this be? If people believed that intoxication would make them too risky in their sexual behaviors, then the more people believed that they were intoxicated the more they might have corrected for this perceived bias. Of course, when people were actually intoxicated, they might have lacked the resources to engage in such corrective actions!

Finally, we note that a lack of an expected subjective experience could also signal to people the existence of a potential bias. For example, consider a situation in which a person tells a joke that he/she obviously believes is funny. If a recipient of that joke expects to be amused but is not, he/she might attempt to identify reasons why the expected reaction was absent (e.g., "perhaps I'm just not in the mood for this type of humor"). At times, such attempts at identifying biasing factors might keep the person from attributing the lack of enjoyment of the joke to the "objective" characteristics of the joke (cf. Gavanski, 1986). Recently, Förster and Strack (in press; Strack & Bless, 1994) used such an idea in studies of recognition memory. Specifically, Förster and Strack (in press) showed that induced theories about how music would inhibit or facilitate memory influenced recognition memory.

12 Of course, if the expectation is based on a theory of bias, then lack of an expected reaction could also signal a lack of bias.
the extent to which people erroneously identified target words as having been presented earlier in the experimental session.

E. INDIVIDUAL DIFFERENCES AND CORRECTION

For the most part, research on corrections has varied features of the situation that make corrections more or less likely (e.g., imposition of cognitive load; subtle prompts to correct). However, as a number of the previous sections illustrated, a variety of individual differences are also likely relevant to theory-based corrections (e.g., see previous section on corrections and stereotyping/prejudice). In fact, a number of individual differences have been investigated within situations relevant to theory-based corrections (although most have been studied within the context of earlier models). As noted earlier, need for cognition was found by Martin et al. (1990) to relate to effects of blatant priming episodes (with people high in need for cognition showing contrast rather than assimilation to blatant primes). When priming is not so blatant, however, individuals high in need for cognition are more likely to show assimilation than those low in need for cognition (see Petty & Jarvis, 1996). When the primes become blatant and therefore more obviously biasing, the default assimilation effect observed for high need for cognition individuals is sometimes overcorrected, resulting in contrast.

Recently, effects of need for cognition have also been studied in attributional contexts where dispositional inferences tend to be the default. People high in need for cognition have been shown to evidence less dispositional attributions than people low in need for cognition (D'Agostino & Fincher-Kiefer, 1992), presumably because high need for cognition individuals are more likely to engage in situational corrections. Individual differences in Need for Cognitive Closure (i.e., a will to have an answer—any answer, Webster & Kruglanski, 1994; but see Neuberg et al., in press) have also been shown to relate to attributional adjustments (e.g., Webster, 1993) and to the impact of priming episodes on impression formation (e.g., Ford & Kruglanski, 1995). Similarly, Personal Need for Structure (M. Thompson, Naccarato, & Parker, 1989; reported in Neuberg & Newsom, 1993) has been shown to influence the impact of priming on person impressions (E. Thompson, Roman, Moskowitz, Chaiken, & Bargh, 1994). Undoubtedly, future work within a theory-based correction perspective will investigate the extent to which these and other observed effects involve theory-based corrections.

It is also likely that other individual differences will be brought to bear on such phenomena, and that moderators of individual-difference-based flexible corrections will be identified. For example, in studies of mood as a biasing factor, people who tend to be quite aware of their mood (Swinkels & Giuli-
ano, 1995), who pay particular attention to their mood (Salovey, Mayer, Goldman, Turvey, & Palfai, 1995), who hold general beliefs that they can regulate moods (Catanzaro & Mearns, 1990; Salovey et al., 1995), or who are high in self-esteem (Rosenberg, 1965; see Smith & Petty, 1995) might be especially likely to identify mood as potentially biasing and might be more likely to engage in theory-based corrections (at least in certain circumstances). In fact, if one looks specifically at the items on the attention subscale of the Trait Meta-Mood Scale (Salovey et al., 1995), many of the items seem to be quite relevant to whether or not people believe that they should avoid using mood in judgments (e.g., "One should never be guided by emotions.").

V. Conclusions

In this chapter we have outlined a flexible correction model that accounts for people's attempts at debiasing their judgments, and we have reviewed research relevant to this framework. We conclude with some comments about the accuracy of judgments resulting from corrections and some suggestions for future research.

A. ACCURACY OF CORRECTIONS

Although people appear to engage in correction processes in a wide variety of social situations with the goal of providing more accurate (or acceptable) judgments, there are numerous instances in which corrections do not result in "debiased" judgments. As documented in this chapter, sometimes this is because people are aware of a bias, but they overcorrect—producing the opposite bias. For example, when correcting for an expected contrast effect, an assimilation bias can be produced (Wegener & Petty, 1995b). When correcting for the perceived negative influence of a disliked source, the disliked source can become more influential than a liked source (Petty et al., in press). At other times, however, "debiased" judgments are not the result because the correction is insufficient (in such a case, the bias is decreased but not eliminated). For example, people might know that a numerical anchor is not correct, but they do not adjust their responses enough to remove the bias (e.g., Tversky & Kahneman, 1974). People might know the stereotype of a social group, but fail to remove fully the effects of the stereotype from judgments of members of the group (Nelson, Biernat, & Manis, 1990). Jurors might know that evidence is inadmissible or people might be asked not to use certain information in impression-formation tasks,
but the information still has an impact (Golding et al., 1990; Thompson et al., 1981). Why do these "failures" of corrections occur?

According to the FCM, there are a number of possible ways in which appropriate corrections might fail to occur. First, people might fail to identify a factor as potentially biasing. If no potential bias is identified, no correction will occur. Even if a factor is identified, people might be unmotivated or unable to engage in corrections. For instance, the object or setting might not be important enough to merit the extra effort necessary to attempt bias removal or the person might view the effect of the factor as legitimate to use in the judgment setting. If the factor is legitimate, little correction is likely and the factor might even have increased influence as motivations for "correct" responses increase.

There are also a number of factors that might influence ability to engage in corrections. For example, people might become less able to correct as cognitive load increases. In fact, research on thought suppression suggests that when people are actively attempting to suppress a thought under cognitive load, this thought can become more accessible than when the thought is not being suppressed (Wegner, 1994; Wegner & Erber, 1992). This suggests that if a person were actively attempting to suppress the perceived effects of some bias on their judgment, the correction would take place under no load (moving judgments away from the perceived effect of the biasing factor), but under load, the bias one is attempting to suppress could have an enhanced impact due to its accessibility (cf. Newman, Duff, Hedberg, & Blitstein, 1996). Thus, lack of ability to engage in theory-based corrections might not only render a correction attempt ineffective, but might actually exacerbate some kinds of bias. Future work might profitably investigate the relations between theory-based corrections and "thought suppression" phenomena.

In addition to cognitive load, people might also have difficulty correcting as the effects of biasing factors are extensively integrated into related knowledge structures and experience. Finally, in some circumstances, peoples' perceptions of bias are incorrect. That is, if people believe that a larger bias is at work than truly exists, overcorrection can be the result. If people underestimate the amount of the bias, undercorrection leaves a smaller version of the bias that initially existed. Interestingly, if people miss the direction of the bias at work, corrections might exacerbate the bias; and, of course, believing a bias exists when it does not can result in theory-based corrections that create biases where initially there were none (e.g., Wegener & Petty, 1995b; see also Wilson & Brekke, 1994).

B. FUTURE RESEARCH DIRECTIONS

Research on theory-based corrections is in the very early stages. Although the research discussed in this chapter provided initial evidence of a number
of important flexibilities in corrections of social judgments, there are a variety of implications of the FCM that await future research. Consider, for example, the notion that motivation and ability to identify potential biases constitute conditions separable from motivation and ability to correct for biases that are identified.

In many cases, factors that impact motivation and ability to identify bias probably also impact motivation and ability to correct. For example, when correction instructions have been used, the instructions directed people toward a factor that might have influenced target assessments and also likely motivated people to remove any biases that were perceived. Likewise, when subtle correction cues have been used (e.g., Petty & Wegener, 1993, study 4), those cues likely prompted people to consider the “first group” of ratings as a potentially biasing factor and prompted people to remove any perceived influence. An extreme form of subtle correction cues would be identified if subliminal priming of corrective goals were found to lead to theory-based corrections. Given that such priming procedures have produced effects similar to more explicit instructions of similar goals (e.g., in studies of impression versus memorization goals; Chartrand & Bargh, 1996), this might be possible. Many variables might similarly impact both the motivation to identify and to correct for perceived biases. For instance, situations that induce feelings of accountability (Tetlock, 1985) or fear of invalidity (Kruglanski & Freund, 1983) might at times increase the likelihood that perceivers will attempt to identify potential biases and to correct for them. Similarly, individual differences in factors such as fear of invalidity might create differences in the likelihood that people will both identify and correct for perceived biases (cf. Martin et al., 1990; Thompson et al., 1994). In fact, future research utilizing manipulations or measures of such constructs would prove useful in more fully understanding the situations in which more naturally occurring events instigate correction processes (e.g., such as serving as a juror versus giving one’s private opinion; Petty et al. 1996).13

Although many variables likely influence both identification of bias and correction for bias, there are likely to be variables that have effects on one rather than the other process. One such factor that we have already discussed as affecting the motivation to correct is the perceived legitimacy of the bias. A person cannot check the legitimacy of a bias unless the potential bias is first identified. Therefore, any factor that influences the perceived legitimacy of a bias can only do so given that people first identify the potential bias. Factors that influence perceptions of legitimacy of bias should

13 Although motivations relevant to a will to be “accurate” might often increase the likelihood of corrections, these motivations are not sufficient for corrections to occur. Recall that corrections only occur to the extent that potential biases are readily identified as such in settings where motivations for accuracy exist. If no biases are identified (or if identified biases are considered to be legitimate—and thus, motivation to correct for them would be low), corrections are unlikely regardless of the amount of overall “accuracy motivation.”
not create differences if the source of bias is not identified. Thus, perceived legitimacy of bias should have its effects primarily on correction rather than identification.

Factors might also impact ability to correct without necessarily affecting ability to identify biases. Consider Martin's (1986) manipulations of completion of the priming task. If, as Martin (1986) reasoned, interruption of a task instigates rumination about the task, then the priming task is likely to be at least as salient to people in interrupted as in completed conditions. Thus, ability to identify the priming task as a source of potential bias should be at least as high in interrupted as in completed conditions. According to Martin (1986), however, rumination might make removal of the influence of the primes difficult (perhaps because of the high accessibility of the activated concepts coloring perceptions of the target information). Thus, ability to correct for the bias introduced by the priming task might be affected more by rumination than would ability to identify the source of potential bias.

Other factors might sometimes relate to both identification and correction but other times relate primarily to correction. For example, variations in general motivation to engage in effortful cognitive activities (i.e., need for cognition, Cacioppo & Petty, 1982) or situational variations that affect motivation without necessarily making the notion of bias salient (e.g., personal relevance, Petty & Cacioppo, 1979; or perhaps personal responsibility, Martin et al., 1990) might influence overall willingness to thoroughly engage in the task at hand (including corrections if biasing factors are easily identified), but might or might not influence the search for potentially biasing factors. This would likely depend on factors such as how perceivers represent the goals and definition of the task at hand.

In addition to distinguishing motivation and ability to identify bias versus correct for it, another of the directions for future research on theory-based correction is to delineate the various correction processes that can be guided by theories of bias. Sometimes theory-based correction is likely to be a relatively simple postjudgment adjustment. At times, however, corrections directed by theories of bias might lead a person to seek out information about the target (either stored in memory or in the judgment environment) that supports a judgment opposed to the person's initial "biased" reaction. In such cases, the "magnitude" of the search for additional information or the extremity of the information that is sought might be tied to the perceived magnitude of the bias at work. In fact, a technique of having people "consider the opposite" (Lord, Lepper, & Preston, 1984) or explain how the opposite might have occurred (Anderson, 1982) has been used to decrease biases associated with perseverance of erroneous beliefs (see also Hirt & Markman, 1995, for effects of a consider-an-alternative strategy). In other circumstances, however, theory-based corrections might proceed through
reinterpretation of the information present at the time of the person's initial reaction (i.e., reinterpretation in light of the "biased" nature of the initial outcome). The processes by which various attempts at correction might take place are also relevant to which factors might influence ability to correct as opposed to ability to identify bias. There are certainly some factors (e.g., existence of varied knowledge about the judgment target) that would be necessary for certain corrections to occur, but would be potentially unrelated to whether or not a source of bias can be identified.

Thus, within the broad class of corrective processes, some corrections might be relatively more effortful or integrative with existing knowledge structures than would others. That is, although corrections might generally take more cognitive effort (at least until a corrective process becomes routinized, cf. Smith, 1989) than simply going with one's initial reaction (with no attempt at correction), some theory-based corrections will likely be more difficult and more thorough than others. To the extent that corrections involve effortful, integrative elaboration of the qualities of targets, those corrected perceptions, like other elaborated judgments, might be more likely to persist over time, resist future attempts at change, and predict behavior better than less well-articulated corrections (see Petty & Cacioppo, 1986; Petty, Haugtvedt, & Smith, 1995; Petty, Priester, & Wegener, 1994).

Additional work is also needed on generation of theories of bias and on ways in which stored theories of bias are modified in specific judgment settings. For example, as our earlier discussion implies, it might be possible to undermine a stored theory of bias (at least in certain circumstances) if reactions to the target seem inconsistent with the theory. For example, consider a situation in which a person possesses a stored theory that a blatant prime of hostility would make a target person seem more hostile than would normally be the case. If, instead of perceiving the target as hostile, however, all of the perceiver's reactions toward the target are suggesting that the target is meek, the perceiver might question the applicability of the theory of bias to this target or situation and the theory would be less likely to guide corrections. Over time, the person might begin to revise the theory or develop beliefs about when the theory is or is not appropriate. It could also be that a well-developed "strong" theory of bias (or a theory that includes numerous settings and/or targets for which the bias occurs) could overwhelm inconsistent reactions to the target and would continue to guide corrections (i.e., the person would reason that the target would seem even more meek if the biasing factors were not present). Such interactions between experiences of the target and theories of how perceptions of the target should be affected constitute an interesting avenue for new research.

Adoption of a theory-based correction perspective also makes additional issues of theory formation and change particularly important. Some biases
and theories about them might be especially stable and long-lived (e.g., biases based on stereotypes of social groups). In much of our initial research, however, the theories we measured regarding how a given judgment context might affect a given target perception are more likely to be short-lived, especially as people gain further experience with that type of context and judgment setting. Therefore, further work in the area of bias correction is likely to profit from research investigating the factors that influence the generation and subsequent perseverance of theories of bias (see Anderson, 1982; Anderson & Sechler, 1986). To a certain extent, factors that influence strength of attitudes (see Petty & Krosnick, 1995) likely also influence the extent to which theories of bias remain stable over time. For example, if a theory of bias is well-practiced (used often), accessible, and well integrated with related knowledge structures, that theory of bias might be relatively likely to persist over time and resist efforts at change. In a practical sense, such research will prove important because efforts at increasing the accuracy of corrections (e.g., in courtroom settings in which accuracy per se is absolutely desirable) and avoiding reverse discrimination (resulting from over-correction) might often become a question of changing theories of bias either in terms of perceived direction or magnitude of bias.

Finally, most of the research on correction processes to date has focused on corrections that were presumably instigated after some initial assessment of a target was made. However, to the extent that a potential bias is noted before target information is encountered, the theory of bias would likely guide initial scrutiny and information gathering regarding the target (or might guide strategic attempts to avoid biasing factors; see also Gilbert, 1993; Wilson & Brekke, 1994; Wilson et al., in press).

C. SUMMARY

In this chapter, the notion of bias correction was reviewed across a variety of research domains. Previous discussions of bias correction were noted, and a view of corrections as guided by perceivers' naive theories of bias (i.e., the Flexible Correction Model; FCM) was outlined. Tests of this perspective have supported critical predictions that differentiate the theory from models of bias correction that focus on "partialling" or subtraction of reactions to contextual stimuli from representations of targets. Also, recent investigations guided by FCM principles demonstrate the relevance of the perspective to a variety of research areas (including persuasion, attribution, impression formation, stereotyping, and mood). Finally, new directions for research were presented. It is our hope that research and theory based on flexible correction notions will help to build a unifying
framework within which correction processes in many areas of psychology can be investigated and explained.

Acknowledgments

Preparation of this chapter was supported in part by National Science Foundation grant SBR 9520854. We thank Russ Fazio, Mark Zanna, and members of the 1991–1996 Ohio State Group for Attitudes and Persuasion (GAP) for comments on this chapter and/or the research reported.

References


Chaiken, S., Liberman, A., & Eagly, A. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), Unin-
tended thought: Limits of awareness, intention, and control (pp. 212–252). New York: Guilford.


Correction for contrast: Thoughtful processes can wipe out or create contrast effects. Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago.


Choose your poison: Effects of lay beliefs about mental processes on attitude change. *Social Cognition*.


Contrast and assimilation with and without awareness: A test of the inclusion/exclusion model. Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago, IL.

