On the role of conflict and control in social cognition: Event-related brain potential investigations

BRUCE D. BARTHOLOW
Department of Psychological Sciences, University of Missouri, Columbia, Missouri, USA

Abstract
Numerous social-cognitive models posit that social behavior largely is driven by links between constructs in long-term memory that automatically become activated when relevant stimuli are encountered. Various response biases have been understood in terms of the influence of such “implicit” processes on behavior. This article reviews event-related potential (ERP) studies investigating the role played by cognitive control and conflict resolution processes in social-cognitive phenomena typically deemed automatic. Neurocognitive responses associated with response activation and conflict often are sensitive to the same stimulus manipulations that produce differential behavioral responses on social-cognitive tasks and that often are attributed to the role of automatic associations. Findings are discussed in the context of an overarching social cognitive neuroscience model in which physiological data are used to constrain social-cognitive theories.

Descriptors: Social cognition, Event-related brain potentials, ERPs, Response conflict, Cognitive control

Scientists who study interpersonal perception and social behavior have long been interested in the role played by covert, rapidly occurring cognitive and affective processes (e.g., Fiske & Taylor, 1991; Markus & Zajonc, 1985). Borrowing largely from paradigms in cognitive psychology, social psychologists have developed clever experimental designs, many involving reaction time (RT) or memory that automatically become activated when relevant stimuli are encountered. Various response biases have been understood in terms of the influence of such “implicit” processes on behavior. This article reviews event-related potential (ERP) studies investigating the role played by cognitive control and conflict resolution processes in social-cognitive phenomena typically deemed automatic. Neurocognitive responses associated with response activation and conflict often are sensitive to the same stimulus manipulations that produce differential behavioral responses on social-cognitive tasks and that often are attributed to the role of automatic associations. Findings are discussed in the context of an overarching social cognitive neuroscience model in which physiological data are used to constrain social-cognitive theories.

This article is based on an address presented upon receipt of the Award for a Distinguished Early Career Contribution to Psychophysiology at the 47th annual meeting of the Society for Psychophysiological Research, Savannah, Georgia, in October 2007. Portions of the research reviewed in this article were supported by research grants from ABMRF/The Foundation for Alcohol Research and the University of Missouri Research Board. Preparation of this article was supported by a grant from the National Institute on Alcohol Abuse and Alcoholism (R21 AA017282).

I thank a number of individuals who have contributed in important ways to the research on which this article is based. First and foremost, I am forever indebted to Gabrielle Gratton and Monica Fabiani for the training in psychophysiology I received from them, for the influence their ideas continue to have on my work, and for setting an excellent example of balancing the demands of academic and family life. I also am grateful to a number of colleagues, especially David Amodio, Eddie Harmon-Jones, Hart Blanton, and Tiffany Ito for their influence on my intellectual development and for their friendship. Finally, the research presented here could not have been accomplished without the help of numerous current and former graduate and undergraduate research assistants, especially Cheryl Dickter, Marc Sestir, Monica Riordan, Sarah Lust, and Erika Henry. Finally, I offer a very special thanks to Susan E. O’Neill for her constant support, friendship, and love.

Address reprint requests to: Bruce D. Bartholow, Department of Psychological Sciences, 210 McAlester Hall, University of Missouri, Columbia, MO 65211, USA. E-mail: BartholowB@missouri.edu

1It is important to stress, however, that the relationship between psychophysiological measures and psychological constructs of interest is complex and rarely (if ever) one-to-one (e.g., Cacioppo & Tassinary, 1990). The point here simply is that augmenting behavioral measures with ERPs or other psychophysiological measures provides a more comprehensive assessment of relevant processes, which can provide a means to more directly test theoretical models of social behavior.
Mechanisms of Social Cognition

A fundamental principle of social cognition is that the beliefs, concepts, and expectancies constituting an individual’s knowledge about the social world are represented in an associative memory network (Carlston & Smith, 1996). This principle has led to the development of theoretical models positing that exposure to a stimulus representing a social category (e.g., a person representing a racial group) spontaneously and effortlessly triggers evaluative and semantic constructs linked to that category (see Macrae & Bodenhausen, 2000; Wheeler & Petty, 2001), leading to facilitated responses (i.e., faster RT) to category-consistent information. Thus, findings in social cognition experiments relying on RT measures often are interpreted in terms of spreading activation (e.g., Collins & Loftus, 1975; Neely, 1977), the idea that cognitive constructs are linked together in memory such that activation of one construct (i.e., via perception of a relevant stimulus) quickly and automatically increases the activation level of other, semantically or affectively related constructs. This basic idea has been used to explain participants’ responses in a large array of so-called priming tasks, including those designed to understand the automatic activation of attitudes (e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986) and activation of stereotypical thoughts and prejudiced feelings about members of racial groups (e.g., Devine, 1989; Dovidio, Evans, & Tyler, 1986; Lepore & Brown, 1997; Fiske & Neuberg, 1990). In such studies, participants’ responses to relevant target stimuli (e.g., attitude words or traits associated with stereotypes) tend to be faster and more accurate when those targets are preceded or accompanied by other stimuli (i.e., primes) that represent a related category (i.e., congruent trials) compared to when primes represent an opposing or unrelated category (i.e., incongruent trials). Such congruency effects generally are attributed to the prime activating a semantic category and, through spreading of activation, constructs related to that category, making them relatively more accessible in memory than less well-related constructs (see Higgins, 1996). Within this framework, priming effects are thought to reflect differences in the strength of automatic associations in long-term memory (e.g., Fazio et al., 1986).

Recently, however, a number of models have been proposed that posit a prominent role for cognitive control in regulating behavioral responses in social cognition (see Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Payne, 2005; Sherman, 2009; Sherman et al., 2008). In general, such models assume that automatic associations are only part of what drives responses to social targets and that the extent to which biases arising from automatically activated associations are expressed behaviorally is determined by the application of controlled, self-regulatory processes. For example, Sherman and colleagues (e.g., Conrey et al., 2005; Sherman et al., 2008) recently have outlined a quadruple-process model (i.e., Quad model) that permits estimation of four distinct processes involved in the regulation of well-learned, prepotent responses in a number of domains: activation, the likelihood that an association is activated when a relevant stimulus is encountered; detection, the likelihood that an accurate or appropriate response to a target can be determined; overcoming bias, the likelihood that an activated association or response tendency can be overcome and replaced with a contextually appropriate one; and guessing, which occurs when no association is activated and a correct response cannot be determined. Applied to the domain of racial stereotyping, the Quad model can be used to understand whether, for example, variability in stereotype-based, biased responses in a given task results from differences in the activation of automatic associations (e.g., Fazio, Jackson, Dunton, & Williams, 1995) versus differences in the ability to overcome prepotent response tendencies based on those associations (see Gonsalkorale, Sherman, & Klauer, 2009).

A core idea underlying effects attributed to overcoming bias is that stereotype-incongruent stimulus situations elicit response conflict, similar to that induced by incongruent trials in the Stroop color-naming task (e.g., MacLeod, 1991). For example, in the Weapons Identification Task (WIT) often used to assess race bias (see Payne, 2001), participants must attempt to quickly categorize (via button press) target objects, presented following White or Black men’s faces, as either handguns or hand tools. Research consistently shows that responses to guns are faster and more accurate following Black compared to White face primes (e.g., Amodio et al., 2004; Amodio, Devine, & Harmon-Jones, 2008; Payne, 2001, 2005; Payne, Shimizu, & Jacoby, 2005). This pattern is thought to occur because exposure to a Black man’s face is associated with a biased tendency to activate the “gun” response, simply on the basis of long-held, stereotypic associations between Black men and violence (e.g., Correll, Park, Judd, & Wittenbrink, 2002; Devine & Elliot, 1995; Sagar & Schofield, 1980). On trials in which the target is actually a tool, this prepotent tendency to respond “gun” conflicts with the correct, goal-driven “tool” response (see also Correll et al., 2002). Thus, although constructs associated with the Black stereotype are thought to become automatically activated upon exposure to a Black face (e.g., Dovidio et al., 1986), overcoming the prepotent tendency to respond “gun” conflicts with the correct, goal-driven “tool” response (see also Harmon-Jones & Winkielman, 2007). ERP Measures of Conflict and Control

A number of ERP components are believed to reflect neural processes of interest for determining the mechanisms that drive social behavior. In particular, two components of the stimulus-
locked ERP have proven useful in our research for investigating the possibility that response conflict occurs in social-cognitive tasks. As described eloquently by Coles (1989; see also De Jong, Wierda, Mulder, & Mulder, 1988; Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988), the lateralized readiness potential (LRP) indexes neural activity in premotor and motor areas of cortex (see Brunia, 1988; Requin, 1985) associated with preparing and generating behavioral responses. For example, as a participant prepares to make a left-hand response, the “readiness potential” (see Kornhuber & Deecke, 1965) will be largest over the right side of the scalp, its amplitude directly reflecting the magnitude of the activated response (see Coles, 1989). The LRP is particularly useful in sequential priming tasks for determining whether and to what extent a response is activated by the prime prior to the onset of the target (see Gratton et al., 1990). Moreover, in tasks involving two response options mapped to opposite hands, the polarity of the LRP reveals which response (e.g., correct or incorrect) is activated by the warning or prime stimulus.

An additional component, the N2 (or N200), consistently has been linked to the hypothesized conflict-monitoring function of the anterior cingulate cortex (see Botvinick, Braver, Barch, Carter, & Cohen, 2001; van Veen & Carter, 2002). The N2 generally has a frontal or fronto-central scalp distribution and is believed to reflect activity in medial-frontal cortical areas (Botvinick et al., 2001; van Veen & Carter, 2002). The N2 tends to be larger on trials involving conflict between competing response representations, such as incongruent Stroop trials (e.g., Liotti, Woldorff, Perez, & Mayberg, 2000) and incompatible flanker trials (e.g., Kopp, Rist, & Muller, 1996). Moreover, N2 amplitude covaries with the degree of incorrect response activation measured via muscle movement (Kopp et al., 1996). The N2 also often is larger on trials requiring a low-frequency response (e.g., Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003), presumably because activating the correct response on such trials conflicts with the (prepotent) response activated by more frequently occurring trials (see Jones, Cho, Nystrom, Cohen, & Braver, 2002).

Although not a measure of response activation or conflict, the parietal P3 (or P300) also can be useful when researchers investigate whether various manipulations that affect response output have effects on stimulus evaluation in addition to or instead of response generation. When elicited by stimuli with evaluative connotations (e.g., positive and negative pictures; see Ito, Larsen, Smith, & Cacioppo, 1998), the P3 is known to be sensitive to the ease with which stimuli can be evaluated categorically: larger P3 amplitude is associated with a greater change in evaluative categorization (see Cacioppo, Crites, Berntson, & Coles, 1993; Ito et al., 1998) and longer P3 latency reflects more effortful evaluation (see Kutatla, McCarthy, & Donchin, 1977; McCarthy & Donchin, 1981). Moreover, the fact that P3 appears to be largely insensitive to response-related processes (see Crites, Cacioppo, Gardner, & Berntson, 1995; Magliero, Bashore, Coles, & Donchin, 1984) means that measuring the P3 in conjunction with components related to response generation (e.g., LRP and N2) provides a way to test the extent to which observed behavioral effects arise from evaluative categorization, response generation, or both (see Gratton, Coles, & Donchin, 1992; Smid, Lamain, Hogeboom, Mulder, & Mulder, 1991; Smulders, Kok, Kenemans, & Bashore, 1995). This issue can be particularly relevant when considering theoretical mechanisms of social behavior tested with reaction-time tasks, such as whether facilitated responses on congruent trials in stereotype-priming tasks reflect more rapid categorization of stereotype-congruent targets, less conflicted response activation, or some combination (see Bartholow & Dickter, 2008; Bartholow, Riordan, Sauls, & Lust, 2009).

Finally, an additional stimulus-locked component, the negative slow wave (NSW), also known as the frontal slow wave because of its typical prominence at frontal and fronto-central scalp locations, can be useful for investigating the extent to which cognitive control processes are brought to bear during social-cognitive tasks. The NSW has been linked to implementation of cognitive control in that it generally is larger on trials in which conflict is successfully resolved, such as incongruent Stroop trials (e.g., West & Alain, 1999; see also Curtin & Fairchild, 2003). In addition, alcohol consumption reduces NSW amplitude (e.g., Bartholow, Dickter, & Sestir, 2006; Curtin & Fairchild, 2003), further implicating this component in cognitive control processes. When used in combination with one another and with behavioral measures and/or in combination with response-locked components such as the error-related negativity (ERN; see Amadio et al., 2004), these components (LRP, N2, P3, and NSW) can shed a great deal of light on the neural mechanisms underlying observed behavioral responses on social-cognitive tasks. Moreover, as the specifics of relevant models are elaborated, additional components might be of use for testing aspects of those models.

Models of Cognitive Control and Behavioral Regulation

Recent years have witnessed a surge of interest in neurocognitive models of behavioral regulation and their application to social behavior. The dual-process model of cognitive control proposed by Botvinick et al. (2001) has been particularly influential. According to this model, an evaluative conflict-monitoring system, subserved by the anterior cingulate cortex (ACC) and nearby medial-frontal regions (see Botvinick et al., 2001; Taylor, Stern, & Gehring, 2007), monitors ongoing behaviors and identifies instances of response conflict or potential conflict (see Carter et al., 1998; Gehring & Fencskai, 2001; van Veer & Carter, 2002). Activity of the conflict monitoring system is thought to manifest at the scalp in the N2 and ERN components of the ERP (see Yeung, Botvinick, & Cohen, 2004). This activity alerts a second, regulatory system, relying on more anterior structures such as dorsolateral prefrontal cortex (dLPFC), which implements top-down control in service of activating an intended response and inhibiting unintended responses (e.g., Botvinick et al., 2001; Carter et al., 2000; Kerns et al., 2004). Implementation of regulatory control is believed to be reflected in the amplitude of the NSW component of the ERP (see Curtin & Fairchild, 2003; West & Alain, 1999, 2000).

A related model, proposed by Holroyd, Coles, and their colleagues (see Holroyd & Coles, 2002), also holds that the ACC is the neural source of the ERN, but ascribes a rather different role.
to the ACC in regulating behavior. Unlike the conflict monitoring function proposed by Botvinick et al. (2001; see also Yeung et al., 2004), Holroyd and Coles posit that the ACC serves as a control filter directly involved in response selection (see also Allport, 1987; Paus, Petrides, Evans, & Meyer, 1993) by learning which of a number of motor controllers is most appropriate for controlling motor output in a given situation. According to their model, the ERN reflects the arrival at the ACC of a signal generated by the mesencephalic dopamine system indicating that action outcomes are worse than anticipated. Thus, whereas the conflict model of Botvinick and colleagues posits that the ERN reflects ACC activation resulting from competing motor responses, the model proposed by Holroyd and Coles explicitly ties the ERN to errors, which serve to tune the learning of appropriate motor control in the ACC.

Numerous theoretical and empirical articles published in the last decade have aimed to support one or the other of these models (e.g., Danielmeier, Wessel, Steinhauser, & Ullsperger, 2009; Scheffers & Coles, 2000; van Veen & Carter, 2002; Ullsperger & von Cramen, 2006) or to challenge them (see Bartholow et al., 2005; Carbonnell & Falkenstein, 2006; Masaki & Segalowitz, 2004), and those will not be reviewed here. For purposes of the current review, it is useful to briefly consider the implications of these models for regulation of social behavior. Perhaps the single most important consideration when applying these or related models to understanding social behavior is the role played by motivational factors. That is, unlike most tasks in which the ACC is elicited and that are used to test theories about its function (e.g., flanker tasks involving letter or arrowhead arrays), social-cognitive tasks typically involve stimuli with inherent motivational significance for participants, such as cues designating social ingroup and outgroup membership or attitude-related words or images. This difference is important given the considerable evidence that the ERN is sensitive to motivational factors—that responses with more “meaning” to participants are associated with larger ERNs (e.g., Amodio et al., 2004; Gehring, Goss, Coles, Meyer, & Donchin, 1993; Gehring & Taylor, 2004; Hajcak, Moser, Yeung, & Simons, 2005).

Although neither model adequately accounts for the role of motivational factors in determining ACC amplitude, predictions concerning motivational significance can be derived from both models. From the perspective of reinforcement learning theory (Holroyd & Coles, 2002), the ERN reflects the ACC learning to map stimuli to appropriate motor controllers. In theory, this process should be modulated by motivational significance (i.e., more meaningful errors should enhance ACC learning). Recent work by Amodio and his colleagues (2004, 2008) largely supports this notion by showing that errors indicating the expression of race bias produce larger ERNs than other errors and that the amplitude of this “race bias” ERN predicts better response control in the task. Moreover, this effect is most pronounced among participants who are internally motivated to avoid expression of race bias (Amodio et al., 2008). The conflict-monitoring theory (Botvinick et al., 2001; Yeung et al., 2004) does not require the assumption that the ACC is modulated by the salience of errors. However, the model does assume that larger ERNs will lead to increased involvement of the dPFC on post-error (or postconflict) trials, indicating attempts to increase response control, a prediction that has been confirmed in at least some studies (see Kerns et al., 2004). If it can be assumed that the size of the ERN reflects (at least in part) the motivational significance of an error, then it follows that greater post-error adjustment should be expected following more meaningful (compared to less meaningful) failures to regulate social responses. Recent work from our laboratory (Bartholow, Henry, Lust, & Sauls, 2009) supports this perspective. However, more work is needed to better understand the implications of motivational factors for both of these models and whether one or the other is more useful for understanding regulation of social behavior.

Racial Categorization, Stereotypes, and Conflict

As reviewed previously, a number of recent studies have conceptualized responses on racial stereotyping tasks as derived from conflict and control processes (e.g., Amodio et al., 2004, 2008; Conrey et al., 2005; Correll et al., 2002; Payne, 2001, 2005; Payne et al., 2005). Evidence supporting this idea has come from studies showing, for example, that the ability to quickly associate Black faces and positive words (as opposed to negative words) relies on overcoming the biased tendency to associate black with negative (Conrey et al., 2005). Similarly, the ability to correctly identify tools following Black face primes in the WIT is positively associated with individual differences in aspects of executive control (Payne, 2005). Also, Klauer and colleagues (e.g., Klauer & Mierke, 2005; Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010) have shown that scores on the Implicit Association Test (IAT; Greenwald et al., 1998) are associated with measures of task-switching, a component of executive control (see Miyake et al., 2000). However, extant studies have not addressed a number of questions about the mechanisms and neural bases of conflict in such tasks, including the stages of processing at which conflict and/ or other operations might occur to influence behavior.

Previous work (see Livingston & Brewer, 2002; Richeson & Trawalter, 2005) has shown that racial categorization is slowed by the presence of context information that is inconsistent with racial stereotypes. Based mainly on principles of spreading activation, this finding has been understood in terms of category membership being more difficult to evaluate when targets are presented in the context of category-inconsistent information (i.e., because activation between race categories and inconsistent concepts takes longer to spread). However, it could be that category-inconsistent information activates response channels opposing those needed to correctly categorize a target individual (i.e., produces response conflict), which also would slow execution of the correct categorization response.

This possibility was investigated in two recent experiments by Bartholow and Dickter (2008). Participants completed a modified Eriksen flanker task (see Eriksen & Eriksen, 1974) in which the target stimuli were men’s faces varying by race (White and Black) and the peripheral flanker stimuli were words associated with stereotypes of Whites and Blacks (see Lepore & Brown, 1997; Wittenbrink, Judd, & Park, 1997). Compatible trials were defined as those in which the race of the target and the flanker words were stereotypically congruent (e.g., Black with “violent”); incompatible trials were those in which target race and flanker words were stereotypically incongruent (e.g., Black with “smart”). Participants were instructed to categorize the target as either White or Black as quickly as possible by pressing one of two buttons and to simply ignore the flanker words (described as distracters). Thus, unlike most traditional flanker tasks in which the flankers themselves are directly linked to a valid response, in this case the flankers were not mapped directly to any response and were only indirectly diagnostic of target categories. We also were interested in whether the relative probability of stereotype-congruent and -incongruent
trials would lead to adjustments in processing strategy similar to that seen in studies using traditional flanker tasks (e.g., Bartholow et al., 2005; Gratton et al., 1992). This was achieved by creating blocks of “expect-compatible” trials (i.e., 80% compatible) and “expect-incompatible” trials (i.e., 20% compatible).

Behavioral data from both experiments were consistent with the notion that stereotype-incongruent flanker words interfered with target categorization, producing a “compatibility effect” (i.e., faster responses on compatible than incompatible trials) similar to that often seen in more traditional flanker tasks (e.g., Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Eriksen & Schultz, 1979; Gratton et al., 1988). Moreover, the compatibility effect was significantly larger in the 80% compatible blocks (M_s = 13 ms and 12 ms in Experiments 1 and 2, respectively) than in the 20% compatible blocks (M_s = 3 ms and 2 ms, respectively), suggesting that participants strategically controlled their attention to the flankers (cf. Gratton et al., 1992). Specifically, when flanker words were likely to be predictive of target race (based on stereotypic associations), participants appeared to attend to and extract information from the entire stimulus array. This was less likely to occur in the 20% compatible condition, in which flanker words were unlikely to provide valid information concerning target race.

The ERP data (Experiment 2) further supported the hypothesis that stereotype-incongruent flankers elicited response conflict. Figure 1 (top panel) shows the LRP (i.e., response activation) elicited by compatible and incompatible trials as a function of block type. In blocks in which compatible trials were more probable, incompatible trials produced a positive deflection in the LRP around 50 ms poststimulus, indicative of a tendency to initially activate the incorrect categorization response prior to activation and execution of the correct response. The presence of this “error dip” on high-conflict trials has been described as evidence of consecutive activation of opposing response tendencies within individual trials (see Gratton et al., 1988, 1992; Spencer & Coles, 1999), in this case indicating that initial response activation was driven by the flankers when compatible trials were more probable (cf. Gratton et al., 1992). The N2 amplitude data (bottom panel of Figure 1) further indicate that conflict was enhanced by the presence of stereotype-incongruent flankers, in that the N2 was enhanced for incompatible relative to compatible trials.

To the extent that the behavioral compatibility effects in this study were related to conflicting response activations on incompatible trials, there should be some association between the size of the compatibility effect in RT and the size of the error dip (essentially, an analogue to the compatibility effect) in the LRP. Correlational analyses supported this prediction, showing that these variables were significantly associated in the 80% compatible blocks (r = .45, p < .05) but not in the 20% compatible
blocks ($r = -0.28, p > .10$). In contrast, differences in RT across compatible and incompatible trials were uncorrelated with P3 latency data ($rs < 0.20, ps > .40$). This lack of association suggests that any potential differences in evaluation difficulty between compatible and incompatible trials were not responsible for observed differences in RT. Moreover, the P3 latency data did not show any compatibility effects ($Fs < 1$), suggesting that evaluative categorization of the stimulus arrays was not affected by flanker compatibility in this paradigm (see also Smid et al., 1991).

Taken together, the data from these experiments support an important role for conflict during response output in determining slowed responses on stereotype-incongruent trials in reaction-time tasks. These findings indicate that preparation of responses can be influenced at a very early stage by peripheral, nontarget information to the extent that an individual expects that information to be useful in aiding appropriate responding to the target. However, relying on such peripheral stimuli produces conflict (and, thus, less efficient responding) when it suggests a response opposed to the one required by the target. Moreover, although connections between the flanks and targets can be attributed to highly accessible, largely automatic associations (i.e., the activation parameter proposed by Sherman et al., 2008), the ERP data from these experiments suggest that the purely associative account implied by a spreading activation model of response facilitation and inhibition (e.g., Fazio et al., 1995) is insufficient to explain the stereotype congruency effects seen here and in other, similar studies (e.g., Amodio et al., 2004; Conrey et al., 2005; Payne, 2001, 2005).

Controlling (and Failing to Control) Race Bias

In recent decades, research on intergroup attitudes has suggested a shift away from negative and toward more positive evaluations of racial outgroups (see Schuman, Steeh, Bobo, & Krysan, 1997). However, despite this trend toward positive overtly expressed racial attitudes, the content of stereotypes about racial minority groups remains quite negative (see Devine & Elliot, 1995; Dovidio & Gaertner, 1998; Fiske, Cuddy, Glick, & Xu, 2002). Research using more indirect methods (e.g., so-called implicit measures) continues to show that minority racial group categories generally are associated with more negative than positive evaluations (e.g., Fazio et al., 1995; Greenwald & Banaji, 1995; Greenwald, Oakes, & Hoffman, 2003; Payne, 2001). This discrepancy between negative, implicit associations and overt expression of positive attitudes suggests that people often make effortful attempts to control the expression of race bias. Such control attempts can be considered a component of a more general skill set associated with self-regulation, which often requires implementation of top-down control over well-learned, prepotent response tendencies in favor of alternative responses considered more appropriate in a given context (e.g., MacDonald, Cohen, Stenger, & Carter, 2000).

Recent work by Amodio and his colleagues (2004, 2006, 2008) has been instrumental in applying aspects of the neurocognitive control framework outlined previously (e.g., Botvinick et al., 2001) to understanding race bias expression and control. These researchers have demonstrated that response errors indicative of race bias are associated with heightened activation of the neural conflict monitoring system, producing enhanced amplitude of the ERN. Moreover, the amplitude of the ERN correlates with estimates of controlled processing on subsequent race bias trials (e.g., more accurate target detection), providing initial evidence of the role of the conflict monitoring system in control of race bias.

A recent experiment by Bartholow et al. (2006, Experiment 2) extended these efforts by testing the role of the regulatory system in race bias control. In particular, this study was aimed at determining how temporary impairment of regulatory control would affect the expression of bias and the neural processes relevant to its control. To accomplish this, Bartholow et al. (2006) had participants consume one of three beverages (a moderate alcohol dose, a higher alcohol dose, or an alcohol placebo) prior to engaging in a go–stop task designed to measure inhibition and expression of bias. The go–stop task is very similar to the stop-signal task, both of which differ from the go/no-go task via the inclusion of “stop” signals on some trials, signaling participants to withhold an activated response. Considerable research indicates that alcohol consumption specifically targets cognitive control resources required for inhibition of prepotent responses, but has virtually no effect on the activation and implementation of responses (see Easdon & Vogel-Sprott, 2000; Fillmore & Vogel-Sprott, 1999, 2000; Mulvihill, Skilling, & Vogel-Sprott, 1997). Moreover, increased inhibition errors (e.g., responding on “stop” trials following alcohol consumption are thought to stem from alcohol’s impairment of cognitive control, rather than from potential effects of the drug on motivation or other aspects of information processing (e.g., Abroms, Fillmore, & Marczinski, 2003). Thus, comparing placebo and alcohol effects provides a relevant and interesting manipulation of the extent to which regulatory control can be implemented when bias control is required.

The task required participants to respond (via button press) as quickly as possible to stereotype-related words following White and Black men’s faces (i.e., primes). On 25% of the trials, however, a visual cue following word onset signaled participants to withhold their response. Trials of interest were those in which the target word was consistent with stereotypes associated with the race of the prime, as these represent prepotent response mappings and therefore should require more regulatory control to successfully inhibit. Therefore, these trials were predicted to be the most difficult for intoxicated participants to manage.

Behavioral data from this study (numbers of inhibition errors on stop trials) are shown in Figure 2A. Analyses of these data showed that whereas errors on stereotype-inconsistent (SI) trials were not significantly affected by beverage type, errors on stereotype-consistent (SC) trials increased significantly as a linear function of alcohol dose. Moreover, only in the highest-dose alcohol condition were inhibition failures more likely on SC than SI trials. These data are consistent with the idea that alcohol specifically impairs regulation of prepotent responses (see also Curtin & Fairchild, 2003), in this case, those associated with the expression of race bias.

Two hypotheses pertaining to the activation of regulatory control were investigated. First, it was hypothesized that if the NSP reflects implementation of cognitive control required for inhibition, the component should be larger on stop trials than on go trials. This hypothesis was confirmed, but only for placebo participants. That is, participants who consumed alcohol showed similar NSP amplitudes on both go and stop trials. Second, to the extent that inhibition of prepotent, race-biased responses requires greater implementation of regulatory control than does inhibition of responses not associated with bias, the NSP should be larger for SC stop trials than for SI stop trials. Here again, however, this prediction applies only to the placebo group. Indeed, as shown in Figure 2B, placebo participants showed larger NSP on SC than SI trials, but high-dose alcohol participants did not.

To the extent that the NSP reflects the implementation of control in order to overcome prepotent response tendencies,
inherent in trying to overcome prepotent, stereotypic associa-
tions by the evaluative, conflict-monitoring system (e.g., Amodio et al., 2004, 2006, 2008; see also Payne et al., 2005). The findings reported by Bartholow et al. (2006) importantly extended this work by demonstrating that successful control over race-biased response tendencies requires intact regulatory control mechanisms, which are significantly impaired by the acute effects of alcohol (see also Casbon, Curtin, Lang, & Patrick, 2003; Curtin & Fairchild, 2003). More recent work in our laboratory (Bartholow, Henry, et al., 2009) indicates that alcohol also can interfere with the conflict monitoring system (see also Ridderinkhof et al., 2002), and that this effect is due, at least in part, to alcohol’s reduction of the distress that typically accompanies the experience of conflict. More generally, findings from this study provide additional evidence of the important role played by cognitive control in social cognition and underscore that a focus on automatic associations cannot fully account for why and under what circumstances biases in overt behavior will be observed.

Response Conflict and Affective Congruency Effects

Although recent models of control in social cognition have focused largely on understanding expression of race bias (e.g., Conrey et al., 2005; Payne, 2005), similar ideas have been debated in the more general area of attitude activation (e.g., Kluger, Burton, & Pitts, 2000), often studied in the laboratory using various affective priming tasks (see Fazio, 2001; Klauer & Musch, 2003). Affective priming (or affective congruency) effects occur when a valenced target stimulus is categorized more quickly as positive or negative (i.e., an evaluative decision task) when it is preceded by an evaluatively congruent prime compared to an evaluatively incongruent prime. First demonstrated by Fazio et al. (1986), this basic finding has been replicated many times (see Fazio, 2001; Klauer & Musch, 2003). Early explanations of the affective congruency effect proposed spreading activation as the likely mechanism, wherein target responses are presumed to be facilitated on congruent trials because evaluative categorization of targets is eased by the prime preactivating the appropriate evaluative construct (e.g., Fazio et al., 1986; see also De Houwer & Hermans, 1994; Hermans, De Houwer, & Eelen, 1994). Some more recent studies similarly have focused on the possibility that affective congruency effects occur at the evaluative categorization stage of processing (see Abrams, Klinger, & Greenwald, 2002; Klauer, Musch, & Eder, 2005).

However, other researchers have begun to conceptualize the effect in terms of conflict at the response output stage of processing. The response conflict model of affective priming holds that both primes and targets activate response tendencies (see Wentura & Rothermund, 2003). On congruent trials primes activate the same response tendency needed to correctly categorize the target, whereas on incongruent trials primes and targets activate opposing response tendencies. Thus, responses on congruent trials are facilitated because the correct target response is partially preactivated by the prime. In contrast, on incongruent trials the response activated by the prime conflicts with the correct target response, thus slowing its execution. Although evidence from several behavioral studies is consistent with this model (e.g., De Houwer, Hermans, Rothermund, & Wentura, 2002; Gawronski, Deutsch, & Seidel, 2005; Klauer & Musch, 2002; Klinger et al., 2000; Wentura, 1999), direct evidence that primes activate response tendencies and that such activation produces conflict on incongruent trials has been lacking.

Graphical data

**Figure 2.** A: Mean numbers of inhibition errors on stereotype-consistent (SC) and stereotype-inconsistent (SI) trials as a function of beverage group (from Bartholow et al., 2006). Whereas errors on SI trials did not differ across beverage groups, errors on SC trials increased linearly with alcohol dose. B: ERP waveforms showing the negative slow wave (NSW) component measured at electrode Cz on successfully inhibited “stop” trials as a function of trial type for those in the placebo and high-dose alcohol groups. Time 0 represents the onset of the stop signal.

NSW amplitude should correlate with the number of inhibition failures on SC trials. Specifically, as the amplitude of the NSW becomes larger (i.e., more negative), inhibition errors on SC trials should decrease, leading to positive correlations. Indeed, this is the pattern we found, with stronger positive correlations at frontal electrodes, particularly left-frontal locations (e.g., \( r = .56, p < .01 \)) and weaker associations at more posterior locations, particularly right-parietal electrodes (e.g., \( r = .17, \text{n.s.} \); see Bartholow et al., 2006, Table 1).

Within the context of the current review, an important question about the findings reported by Bartholow et al. (2006) is whether alcohol effects on expression of bias reflect differences in the activation of automatic associations, the ability to regulate responses based on those associations, or some combination. Although Bartholow et al. (2006) argued that their findings reflected effects on regulation of biased responding, Sherman et al. (2008) recently addressed this question more directly by analyzing the inhibition error data from this study using formulas in their Quad model. Their analyses indicated that, compared to the placebo beverage, the alcohol beverage impaired the parameter associated with overcoming bias, but had no effect on the activation, detection, or guessing parameters. Thus, consistent with the argument put forth by Bartholow et al. (2006), these findings indicate that alcohol has no effect on the activation of stereotypic associations, but rather interferes with the ability to regulate relevant responses once those associations are activated.

Findings from other, previous studies have demonstrated that race-biased responding can occur despite the detection of conflict inherent in trying to overcome prepotent, stereotypic associa-
In a recent experiment, Bartholow, Riordan, et al. (2009) used ERPs to directly test the hypothesis that incongruent trials in affective priming elicit response conflict. This hypothesis rests on two assumptions: (1) that primes activate response channels prior to target onset and (2) that this activation generates conflict on incongruent relative to congruent trials. A further assumption is that this conflict slows response execution on incongruent trials. To test these assumptions, we measured the LRP and N2 components while participants completed an evaluative decision task. Any preferential response activation following prime onset (and prior to target onset) would be evident in the amplitude and polarity of the LRP. To the extent that such response activation conflicts with the response required by a target, the N2 should be enhanced following target onset.

In addition, to investigate the extent to which participants can strategically control their responses, we manipulated the proportion of congruent and incongruent trials across trial blocks (see also Klauer, Rossnagel, & Musch, 1997; Spruyt, De Houwer, Hermans, & Eelen, 2007). To the extent that response facilitation on congruent relative to incongruent trials is driven by automatic spreading of activation, congruent trial responses should be faster than incongruent trial responses regardless of contextual factors such as relative probability (see Spruyt et al., 2007). In contrast, strategic control of response activation would be evident by differential behavioral responses—and patterns of response activation and conflict in the ERP—when the probability of congruent trials is relatively low compared to when it is high (see Bartholow et al., 2005; Gratton et al., 1992).

The behavioral data from this experiment produced significant Congruence × Probability interactions for both reaction times and error rates, replicating previous studies showing that the size of the affective congruency effect varies along with the proportion of congruent trials (e.g., Klauer et al., 1997; Spruyt et al., 2007). Of greater interest here, the LRP and N2 amplitude data also were modulated by congruence probability in a manner consistent with the response conflict hypothesis. As shown in Figure 3 (left panel), when congruent trials were highly probable participants began to preferentially activate the congruent target response before the target appeared (e.g., activating the “positive” response to positive primes), as evinced by the relatively negative amplitude of the LRP for congruent trials compared to incongruent trials. Due to this preferential response activation conflicts with the response required by a target, the N2 should be enhanced following target onset.

Moreover, covariance analyses showed that variation in behavioral responses (RT) was dependent on variations in both LRP and N2 amplitude. For example, when the RT data were

Figure 3. ERP waveforms showing the lateralized readiness potential (LRP) measured from C3 and C4 electrodes (left panel) and the N2 component measured at FCz (right panel) as a function of target congruence and the probability of congruent trials (from Bartholow, Riordan, et al., 2009). Vertical marks on the LRP waveform labeled “P” and “T” represent the time of prime and target onset, respectively. Of primary interest here was the amplitude of the LRP between prime onset and target onset, which indicates relative response activation elicited by the primes. The vertical arrow (and “T”) on the timeline for the N2 represent target onset. Conflict monitoring reflected in the N2 increases as a function of the extent to which primes preferentially activate the response opposite the one required by the target.
Conclusions

The findings of the studies reviewed in this article point to an important role for cognitive control and response conflict resolution in understanding behavioral responses typically observed in a number of laboratory paradigms used to study social cognition. More generally, the data from these studies are consistent with a number of recent models in which self-regulatory control processes assume a prominent role in explaining social behaviors, particularly in the domains of racial stereotyping and evaluations/attitudes (e.g., Conrey et al., 2005; Payne, 2005; Sherman et al., 2008). Perhaps the primary importance of these and related findings (e.g., Klauer & Teige-Mocigema, 2007; Radvansky et al., 2008; Teige-Mocigema & Klauer, 2007; von Hippel, 2007) lies in their implications for understanding individual differences in the expression of many forms of bias in terms of the ability and/or motivation to exert control over one’s responses rather than differences in the strength of automatic associations, a prominent feature of many previous models (e.g., Fazio et al., 1986, 1995; Greenwald et al., 1998).

Beyond their implications for recent models of bias expression, the findings reviewed here more generally underscore the usefulness of physiological responses for constraining theories about cognitive and social-cognitive processes (cf. Mangum & Hillyard, 1995). The psychophysiological measures used in these and other studies (for reviews see Bartholow & Amadio, 2009; Bartholow & Dickter, 2007, 2010) provide a means to covertly assess neurocognitive responses with known links to psychological processes, allowing researchers to determine whether stimulus conditions of interest evoke differences in the involvement of those processes. Even with carefully planned experimental designs, it often is difficult to separate the contribution of multiple, sometimes overlapping processes to behavioral outcomes with behavioral measures alone (see Coles et al., 1995). However, augmenting behavioral response measures with ERPs (and/or other measures of physiological response) provides an opportunity to more comprehensively assess the mental and sensory operations that lead to theoretically relevant differences in human behavior.

As the field of social neuroscience continues to develop over the coming years, it will be important for researchers to consider the usefulness of multiple measures for understanding the underlying neural events that give rise to social behavior. In particular, at present there appears to be a bias within the field favoring the contributions made by functional magnetic resonance imaging (fMRI) over electrophysiological methods and a lack of understanding among some researchers that ERP and electroencephalogram measures provide more direct assessment of neural activity than fMRI or positron emission tomography. Clearly, fMRI and related methods provide critical information concerning the role of specific neural structures in producing social behavior. However, as I hope is clear from the current review, ERP measures can be just as important in constraining social-cognitive theories and, in fact, are more informative in terms of the timing of neural responses than is fMRI. Future work should strive to exploit the benefits of multiple measures (e.g., ERP, fMRI, and event-related optical signals) in order to best understand the complex spatial and temporal relationships within the brain that determine and modulate social behaviors.

REFERENCES


(Received September 8, 2009; Accepted October 13, 2009)

B. D. Bartholow

Harvard University Press.