Article

A Functional Coupling of Brain and Behavior During Social Categorization of Faces

Personality and Social
Psychology Bulletin
2021, Vol. 47(11) 1580–1595

DOI: 10.1177/0146167220976688
journals.sagepub.com/home/pspb
SAGE

Hannah I. Volpert-Esmond[1] and Bruce D. Bartholow[2]

Abstract

Considerable research has focused on how people derive information about others' social category memberships from their faces. Theoretical models posit that early extraction of task-relevant information from a face should determine the efficiency with which that face is categorized, but evidence supporting this idea has been elusive. Here, we used a novel trial-level data analytic approach to examine the relationship between two event-related potential components—the P2, indexing early attention to category-relevant information, and the P3, indexing stimulus evaluation—and the speed of overt categorization judgments. As predicted, a larger face-elicited P2 on a particular trial was associated with faster overt race or gender categorization of that face. Moreover, this association was mediated by P3 latency, indicating that extraction of more category-relevant information early in processing facilitated stimulus evaluation. These findings support continuous flow models of information processing and the long-theorized functional significance of face-elicited neurophysiological responses for social categorization.

Keywords

person construal, event-related potentials, social categorization, brain–behavior relationships, continuous flow

Received May 26, 2020; revision accepted November 5, 2020

Person construal—the process of placing people into meaningful social categories such as race and gender—is a fundamental component of human cognition (Bodenhausen et al., 2012; Macrae & Bodenhausen, 2000). Face processing plays an enormous role in person construal (Adolphs, 2001; Kramer et al., 2017; Macrae et al., 2005; Mason & Macrae, 2004). From their faces, people's social category memberships are determined within mere hundreds of milliseconds (Freeman et al., 2010; Ito & Bartholow, 2009; Rule & Ambady, 2010). A prominent theory of person construal known as the dynamic interactive (DI) model (Freeman & Ambady, 2011, 2014; Freeman & Johnson, 2016) posits that person construal is the outcome of a dynamic integration of "bottom-up" visual processing of facial features (e.g., physiognomy, skin tone) with higher order, "top-down" information-processing operations (e.g., prior knowledge, current goals). This integration is posited to unfold over iterative cycles until a stable representation of a target's social category is achieved. In theory, processes contributing to these iterations should be reflected in face-elicited neurophysiological responses (Freeman & Johnson, 2016; Stolier & Freeman, 2017) and variability in those responses should contribute to variability in overt categorization judgments. The very rapid time course over which these posited iterations are thought to cycle—on the order of tens of milliseconds (Freeman & Johnson, 2016)—makes them very difficult to observe with temporally sluggish hemodynamic functional brain imaging (i.e., functional magnetic resonance imaging [fMRI]). Here, we used event-related (brain) potentials (ERPs), which can capture face-elicited neural responses with millisecond-level temporal precision (Amodio et al., 2014; Levsen et al., 2018), to track the rapidly occurring neural responses elicited by faces and assess their functional significance for race and gender categorization judgments.

Information Processing and Person Construal

According to the DI model, initial perception of a face simultaneously activates multiple possible social category representations. A dynamic competition among these possible representations then ensues in which bottom-up cues gleaned from the face—many of which are consistent with multiple

---

[1] The University of Texas at El Paso, USA
[2] University of Missouri, Columbia, USA

Corresponding Author:
Hannah I. Volpert-Esmond, Department of Psychology, The University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968, USA.
Email: hivolpertes@utep.edu

possible categories—are integrated with top-down knowledge and the perceiver's goals. This model is grounded in recurrent connectionist network models of information processing (McClelland, 1991; McClelland & Rumelhart, 1981), which emphasize the importance of context and existing knowledge for determining how visual percepts are understood (also see Bruner, 1957). Evidence supporting the idea of parallel activation of multiple categories during person construal comes from studies showing that competing categorizations (e.g., male vs. female) elicit conflicting motor output, as reflected in measures of computer mouse tracking (Freeman, Penner et al., 2011; Freeman et al., 2016; Hehman et al., 2014) and motor cortex activation (Freeman, Ambady, et al., 2011), which, in turn, affects the time needed to make an overt categorization response.

The DI model bears a number of striking similarities to earlier continuous flow (CF) models of information processing, not specific to faces (Coles et al., 1985, 1995; Eriksen & Schultz, 1979; Requin et al., 1988). According to the CF hypothesis, initial information extracted from a stimulus is consistent with multiple possible identifications, particularly when features of the context create ambiguity. As information about stimulus identity accumulates, activation potential simultaneously accumulates in viable response channels; an overt response is evoked when activation in a given channel exceeds some criteria. Some of the first direct support for this conception was provided by Coles and colleagues (1985), who used ERPs to show that partial activation of multiple behavioral responses is evident in motor cortex several hundred milliseconds before stimulus evaluation has been completed—as reflected in the latency of the P3 component of the ERP—and before an overt response is emitted (also see Gratton et al., 1988). Such findings support the CF model's assumption that stimuli are evaluated gradually, and that stimulus evaluation and response preparation unfold in parallel (Coles et al., 1995).

Of particular relevance to the current research, both the CF and DI models assume that information about the stimulus accumulates dynamically over time. Accordingly, the amount of task-relevant information extracted early in processing should determine the efficiency of stimulus evaluation and overt classification (Coles et al., 1995). Research on the processing of priming or warning stimuli provides an example (see Gratton et al., 1990; Hackley, 2009). As reported by Gratton et al. (1990), relative to priming stimuli that did not predict the subsequent target's identity, priming stimuli that predicted the identity of an upcoming target stimulus elicited larger amplitude of the P3 ERP component (also see Bartholow et al., 2009), reflecting the extraction of more evaluative, decision-relevant information from the primes (Duncan-Johnson & Donchin, 1982; Polich et al., 1996; Nieuwenhuis et al., 2005). Moreover, larger prime-elicited P3 amplitudes predicted faster correct target classifications, consistent with voluminous research showing that extracting information about the identity of a stimulus early in processing facilitates stimulus evaluation and identification (Bargh, 2014; Hackley, 2009; Woodworth, 1938).

This same principle should apply to social categorization from faces: Extraction of more category-relevant information early in processing should facilitate evaluation and overt identification of the social category or categories that the face represents. Researchers have identified an early-latency (~180 ms following face onset) deflection in the ERP waveform known as the P2 that appears to reflect processes associated with the attempt to extract information about a target person's social category memberships. The face-elicited P2 is highly sensitive to social categories, such as race and gender (Amodio, 2009; Correll et al., 2006; Dickter & Bartholow, 2007; Dickter & Kittel, 2012; He et al., 2009; Ito & Bartholow, 2009; Ito & Tomelleri, 2017; Ito & Urland, 2003, 2005; Kubota & Ito, 2007, 2017; Simon & Gutsell, 2019; Volpert-Esmond & Bartholow, 2019; Volpert-Esmond et al., 2017; Willadsen-Jensen & Ito, 2006, 2008, 2015). In the broader information-processing literature, P2 amplitude has been associated with allocation of visual selective attention to goal-relevant stimuli (Hillyard & Münte, 1984; Luck & Hillyard, 1994; Wijers et al., 1989), which improves perception and performance by modifying sensory inputs early in processing (Mangun, 1995). Thus, it stands to reason that the face-elicited P2 reflects allocation of attention to features of faces that enhance perception and facilitate identification of social categories. Supporting this idea, recent findings show that top-down categorization goals enhance the influence of goal-relevant features and suppress the influence of goal-irrelevant features on face-elicited P2 amplitude (Volpert-Esmond & Bartholow, 2019).

Associating Brain and Behavioral Responses in Person Construal

Despite the apparent relevance of the P2 for early extraction of social category information during person construal, research to date has failed to demonstrate any functional association between face-elicited P2 amplitude and the efficiency of social categorization. Two previous studies have tested the hypothesis that a larger P2 will be associated with faster categorization reaction times (RTs; Dickter & Bartholow, 2007; Kubota & Ito, 2007); neither found support for this hypothesis. Critically, however, in both of these studies, the traditional signal-averaging approach was used (Woodman, 2010), wherein the ERP signal is averaged across all trials within each experimental condition, producing a single observation per condition for each participant. Although signal averaging increases the otherwise low signal-to-noise ratio inherent in neurophysiological responses (Dawson, 1954; Logothetis & Wandell, 2004; Luck, 2014), this approach severely limits the extent to which ERPs can provide information on the processing of a given stimulus or event in a particular instance, as opposed to the average trend across the task (see Coles & Rugg, 1995). Moreover, the signal averaging approach rests on a number of assumptions that, on their face, are implausible. First, use of signal averaging assumes that the neurophysiological signal of interest is stable over repeated assessments, such that the average across trials accurately represents any particular response. Previous work (Volpert-Esmond et al., 2017, 2018; Von Gunten et al., 2018) has demonstrated the inadequacy of this assumption, showing that the amplitude of a number of well-studied ERPs, including the face-elicited P2, changes over the course of a laboratory task (also see Berry et al., 2019; Brush et al., 2018; Regtvoort et al., 2006).

A second assumption of the signal-averaging approach is that the only meaningful variability in the ERP response of interest occurs between rather than within persons. With signal averaging, any within-person variability in the response across trials is assumed to be noise and therefore is disregarded. Previous research points to the inadequacy of this assumption too. For example, trial-to-trial variation in the latency of the P3 component, known to reflect the speed of stimulus evaluation (Kutas et al., 1977), predicts RT to overtly classify the eliciting stimulus (e.g., Coles et al., 1985; McCarthy & Donchin, 1981; Sassenhagen et al., 2014).

Examining within-person relationships between brain and behavior requires modeling responses at the level of individual trials, rather than averaging across trials (Pernet et al., 2011). Recently, multilevel modeling (MLM) has been proposed as a useful approach for this purpose (e.g., Tibon & Levy, 2015; Volpert-Esmond et al., 2018; Vossen et al., 2011). Through the use of nesting and modeling of random effects, MLM can account for a number of unique sources of error variability, including individual participants and electrode channels, thereby partitioning more sources of variance from the error term and increasing power to detect fixed effects (Gelman & Hill, 2007; Vossen et al., 2011). These features allow researchers to utilize trial-level data without first reducing noise through signal averaging.

Of particular importance for our purposes, MLM additionally can be used to separate between- and within-person sources of variability in trial-level relationships among variables—something that has not been possible using alternative trial-level ERP approaches (Coles et al., 1985; Debener et al., 2005; Philiastides et al., 2006; Ratcliff et al., 2009; Sassenhagen et al., 2014). Dissociating these sources of variability has proven important in previous work examining processes that, in theory, should be expected to vary more within than between individuals. In one example, Von Gunten et al. (2018) tested the between- and within-person relationships between ERPs reflecting cognitive control and RTs during a response conflict task. Their results showed that, on a given trial, control-related ERP responses that were relatively large (compared with an individual's mean ERP response) predicted faster correct responses to targets on the next trial. However, no between-person relationships were observed; that is, a person's average control-related neurophysiological response did not predict that person's response behavior relative to other people. This could reflect the fact that associations within individuals can vanish or even reverse when individuals' data are combined (i.e., Simpson's Paradox; Simpson, 1951; Wagner, 1982). Thus, the ability to appropriately disaggregate within- and between-person sources of variability appears critical to any effort to associate neural responses with behavior.

The Current Research

A primary aim of the current research was to demonstrate the functional significance of the face-elicited P2 for person construal by testing whether within-person variability in P2 amplitude across trials predicts RT for overt categorization decisions. Based on the notion that face-elicited P2 amplitude reflects allocation of attention to features that facilitate social category distinctions (Amodio, 2009; Correll et al., 2006; Ito & Bartholow, 2009; Kubota & Ito, 2007), and the idea that greater extraction of information early in processing should facilitate overt categorization judgments (Freeman & Ambady, 2014), we predicted that larger P2 amplitudes would facilitate overt classification of faces. In addition, we examined how top-down task goals influence early extraction of information by separately examining the relationship between P2 amplitude and RT when participants categorized faces by race or by gender.

To further test the idea that P2 amplitude reflects extraction of information helpful for evaluating social categories, we also tested whether P2 amplitude is negatively associated with P3 latency on individual trials. Nearly a half-century of research supports the conclusion that the latency of the P3 reflects the time required to evaluate the eliciting stimulus (see Coles et al., 1990; Donchin, 1979; Donchin & Coles, 1988; Polich, 2012). Specifically, the latency of the P3 places an upper limit on stimulus evaluation time, meaning that by the time the P3 peaks in amplitude, the processes required to categorize the stimulus must have occurred. This conclusion is supported by the observation that as stimulus categorization becomes more difficult, P3 latency increases (Kutas et al., 1977; McCarthy & Donchin, 1981). Moreover, whereas overt responses are the product of numerous psychological (e.g., stimulus evaluation, response competition) and kinesiological (e.g., motor speed) processes, P3 latency is considered a purer measure of the duration of stimulus evaluation that is not confounded by those other factors (Coles et al., 1995). Here, we predicted that trials in which the P2 is relatively large would also elicit a relatively fast P3 latency (i.e., P2 amplitude would negatively predict P3 latency). Furthermore, we predicted that P3 latency would mediate the predicted association between P2 amplitude and categorization RT on individual trials—further supporting the idea that extraction of more category-relevant information early in processing will facilitate evaluation of a face, thereby speeding overt classification. These predicted associations were expected to manifest in within-person effects but not between-person effects.

General Method

Studies 1 and 2 report new data from samples previously reported in Volpert-Esmond et al. (2017) and Volpert-Esmond and Bartholow (2019), respectively. Here, we report only data from the face-categorization tasks administered in each study. Methodological details for additional tasks and questionnaires administered in those first two studies can be found in those previous reports. Study 3 reports data from a novel sample. Samples from all three studies were collected at a predominantly White institution in the Midwest.

Measures

In all studies, participants completed a race-categorization task in which they viewed Black and White faces with neutral expressions. In Studies 2 and 3, participants also completed a gender-categorization task where they viewed the same faces and categorized them by gender. The order of the race- and-gender categorization tasks in Studies 2 and 3 were randomized across participants. Electroencephalogram (EEG) was recorded during all tasks, and stimulus presentation and behavioral response recording was controlled using E-Prime (Psychology Software Tools, Inc., Sharpsburg, PA). Face stimuli were taken from the Chicago Face Database (Ma et al., 2015) and were converted to gray scale to reduce differences in low-level perceptual features across faces. In addition, the brightness and contrast of the images were adjusted to be equivalent across stimulus conditions, using the SHINEd toolkit in MATLAB (Willenbockel et al., 2010). There were slight methodological differences across studies, which were a function of the primary theoretical purposes of each original study. Specifically, Studies 1 and 2 incorporated a fixation manipulation, such that faces were presented in one of two locations. In one location, the fixation cross preceded the middle of the forehead (i.e., forehead fixation) and in the other location, the fixation cross preceded the nasion (between the eyes, that is, eye fixation).[1] In Study 3, participants fixated only between the eyes. Due to the inclusion of the fixation manipulation and/or female faces, the number of times each unique face stimulus was presented varied across studies to maintain similar numbers of total trials and reduce participant fatigue. In Study 1, 32 faces of each race (32 Black male faces, 32 White male faces) were presented 4 times each, resulting in 128 trials in each race category and 256 trials total in the race-categorization task. In Study 2, 32 faces of each race and gender (32 Black male faces, 32 Black female faces, 32 White male faces, and 32 White female faces) were presented 2 times each in each task, resulting in 64 trials in each race/gender category and 256 trials total in each task. Study 3 used the same stimuli as Study 2, but faces were presented only once in each task, resulting in 32 trials in each race/gender category and 128 total trials in each task.

Trial structure was the same in all studies and tasks. On each trial, a fixation cross was presented (jittered: 500, 700, or 900 ms in Study 1; 500, 650, or 800 ms in Studies 2 and 3), followed by a face (270 ms), which was then followed by a visual mask (530 ms). Participants were instructed to categorize the face by race (race-categorization task) or gender (gender-categorization task) as quickly and accurately as possible, using one of two buttons on a game controller. Failure to respond within 800 ms following face onset elicited a "TOO SLOW" warning displayed for 1,000 ms. The ITI was 600 ms.

Electrophysiological Recording and Data Processing

In all studies, the EEG was recorded using tin electrodes embedded in a stretch-Lycra cap (Electro-Cap International, Eaton, OH) although the number of electrodes varied across studies (20 in Study 1, 35 in Study 2, and 33 in Study 3).[2] Scalp electrodes were referenced online to the right mastoid; an average mastoid reference was derived offline. Signals were amplified with a Neuroscan SynAmps amplifier (Compumedics, Inc., Charlotte, NC). Online filtering and sampling parameters varied slightly across studies (Study 1: 0.10–30 Hz[3] bandpass at a sampling rate of 1000 Hz; Studies 2 and 3: 0.05–40 Hz[4] bandpass at a sampling rate of 500 Hz). Impedances were kept below 10 KΩ. Ocular artifacts (i.e., blinks) were corrected from the EEG signal offline, using a regression-based procedure (Semlitsch et al., 1986). Trials containing voltage deflections greater than ±75 microvolts (μV) were discarded.

P2 quantification (amplitude). In all studies, visual inspection of the face-elicited waveforms (see Figure 1) showed a positive-going deflection prominent over central and centro-parietal scalp locations and peaking roughly 160 ms after face onset, consistent with previous characterizations of the P2 during face processing (Dickter & Bartholow, 2007; Ito & Urland, 2005). In all studies, the P2 was quantified in each trial as the mean amplitude from 130 to 190 ms post-face onset at seven central and centro-parietal locations (Cz, C3, C4, CPz, CP3, CP4, and Pz). Coverage was extended in Study 2 to include C1, C2, CP1, and CP2, and in Study 3 to include C1 and C2. In Study 1, the median number of trials included for each type of stimulus (Black men/White men) for each participant was 114 (min 66, max 126). In Study 2, the median number of trials included for each type of stimulus (Black/White men/women) for each participant in the race- and gender-categorization tasks was 54 (min 26, max 64) and 51.5 (min 25, max 64), respectively. In Study 3, the median number of trials included for each type of stimulus (Black/White men/women) for each participant in the race- and gender-categorization tasks was 28 (min 11, max 32) and 29 (min 13, max 32), respectively.

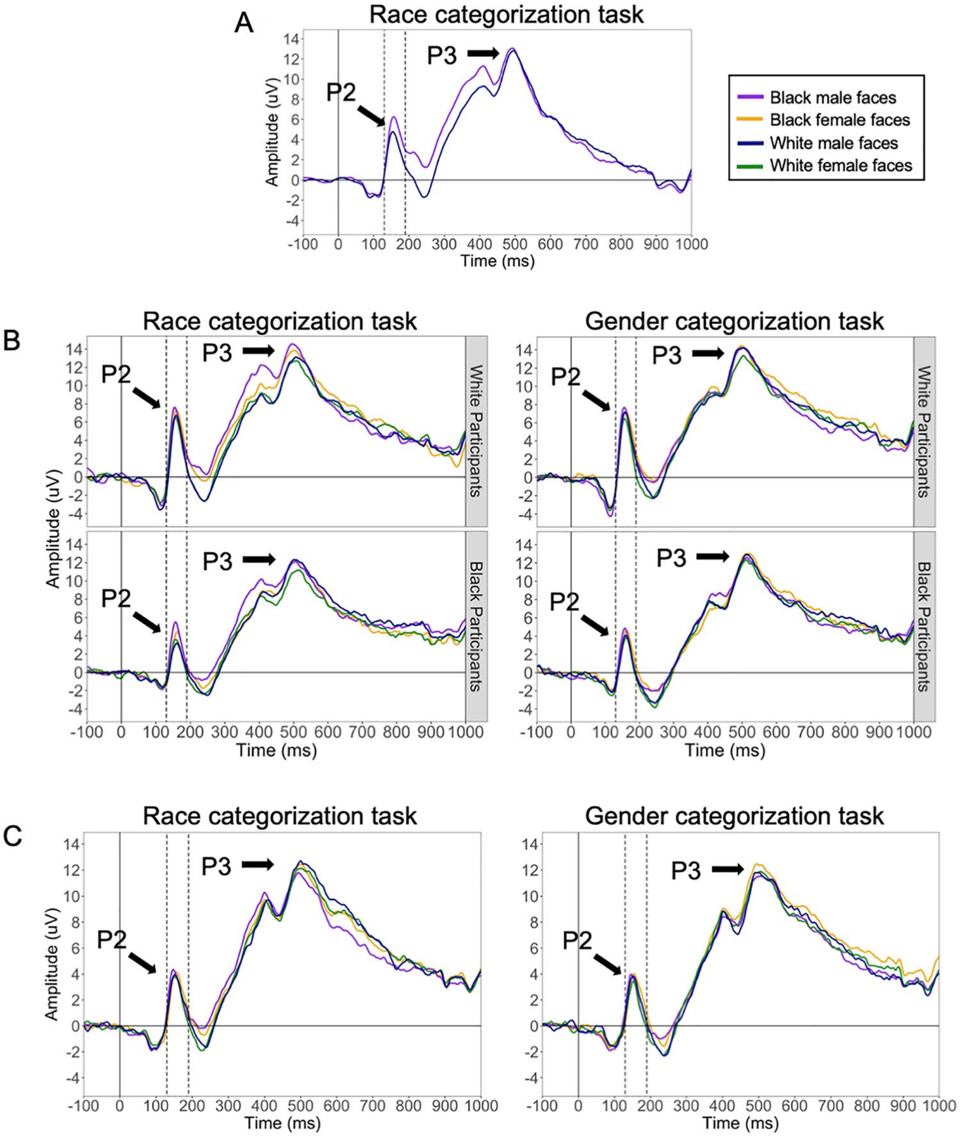

Figure 1. Grand average waveforms recorded at electrode CPz (Study 1 in Panel A, Study 2 in Panel B, and Study 3 in Panel C).
Note. In Study 1, all faces presented were male (Black/White). In Study 2, male and female faces were presented (Black/White). To increase visibility, only waveforms where faces were presented in the eye fixation condition are presented (Studies 1 and 2). However, all data (including forehead fixation trials) were included in the analyses. Dashed vertical lines depict the time window during which mean P2 amplitude was quantified (130–190 ms post-face onset).

P3 quantification (latency). Visual inspection of the face-elicited waveforms (see Figure 1) showed a positive-going deflection, peaking roughly 500 ms after face onset and most prominent over central and parietal scalp locations. Because of the large variability in the latency of the P3 across participants, each individual's average P3 peak latency was first determined from an individual grand-averaged waveform, including all trials and conditions. In Studies 2 and 3, each individual's average P3 peak latency was assessed separately in the race-categorization task and gender-categorization task. Thereafter, the peak latency of the P3 component in each trial for each individual was determined as the peak latency within a 250 ms window surrounding that individual's average peak latency. P3 latency was quantified at the same electrodes as P2 amplitude in each study.

Statistical Approach

We used the R package lme4 (Bates et al., 2015) to fit multilevel models for data analysis. We were primarily interested in the trial-level relationship between the P2 amplitude elicited by a face and the speed with which participants made relevant categorization responses. Thus, we used RT on each trial as the dependent outcome. To separately estimate the contribution of between-person and within-person variability in P2 amplitude to RTs, we used the disaggregation approach described by Curran and Bauer (2011). In the current context, this involved separating the P2 amplitude variable into two separate predictors. The first predictor—each participants' mean P2 amplitude—was entered as a Level-2 (person level) predictor and represents the between-person effect. The second predictor—the P2 amplitude for each trial, centered around the participant mean—was entered as a Level-1 (trial level) predictor and represents the within-person effect. Creating the participant mean and centering trial-level data around that mean was done separately for each electrode, resulting in an intercept of 0 for each electrode. Therefore, electrode was not included as a random effect in the models.[5]

In addition, we tested P3 latency as a mediator of the relationship between P2 amplitude and RT. To test for the presence of mediation in each study, we fit two separate multilevel models, using the multilevel mediation approach reviewed in Preacher (2015). The first model estimates the effect of P2 amplitude (X) on P3 latency (M): the *a* path.[6] The second model estimates the joint effect of P2 amplitude (X) and P3 latency (M) on RT (Y): the *b* and *c'* paths.[7] Thereafter, we calculated the indirect effect by multiplying the *a* and *b* paths and used the R package RMediation (Tofighi & MacKinnon, 2011) to produce 95% confidence intervals (CIs) for the indirect effect. Specifically, we used the asymmetric confidence limits approach with modified critical values from the empirical-M method (MacKinnon et al., 2004), which has been advocated as the most appropriate approach to produce CIs for the indirect effect in multilevel mediation (Pituch et al., 2005; Pituch & Stapleton, 2008; Preacher, 2015).

Importantly, as the X, M, and Y variables are measured at Level 1 (described as a 1-1-1 mediation model), these models—as typically specified—confound between- and within-person relationships. Thus, we applied the same disaggregation approach (Curran & Bauer, 2011) to the mediation models, termed the *unconflated multilevel model* approach to mediation (Preacher et al., 2010; Zhang et al., 2009). In the two models testing the *a* path and the *b* path, separate predictors capturing the between- and within-person variability in X and M were included in the models. By separately estimating the between- and within-person effects in each path, we can separately estimate the within-person and between-person indirect effect. Creating the participant mean and centering trial-level data around that mean was done separately for P2 amplitude and P3 latency at each electrode within arrays used for analyses, resulting in an intercept of 0 for each electrode. Therefore, electrode was not included as a random effect in the models.

For all models, Satterthwaite approximations were used to estimate degrees of freedom and to obtain two-tailed *p* values; in situations where degrees of freedom were >200, we report the results as *z* statistics. For models testing the relationships between P2 amplitude, P3 latency, and RT, the intercept was allowed to vary by participant but no random slopes were included. Data and all code used for analysis can be found at https://github.com/hivolpertes/WithinSubP2.

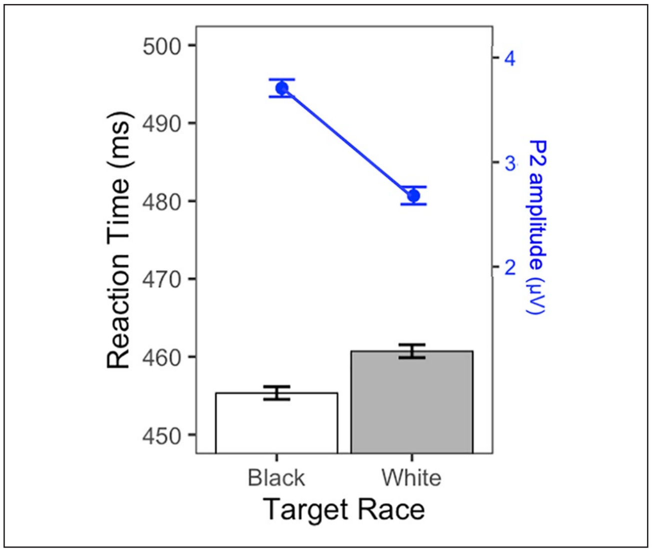

Figure 2. Mean RTs and mean P2 amplitude by stimulus condition in Study 1 (race-categorization task).
Note. White and gray bars correspond with *y*-axis on the left (RT) and blue points correspond with *y*-axis on the right (P2 amplitude). Error bars represent 95% CIs. *N* = 62, roughly 53% women, 90% White/non-Hispanic. RT = reaction time; CI = confidence interval.

Study 1

Participants

Sixty-five individuals (34 women, 31 men) participated in exchange for credit toward a research requirement in an introductory psychology course or for monetary compensation. Participants ranged in age from 18 to 48 years (*M* = 20.4). Sixty self-identified as White, two identified as Asian, and three identified as more than one race. None identified as African American. Because multilevel models can accommodate low numbers of trials per condition, participants were not excluded based on number of trials per condition, as is often done (Luck, 2014). However, due to concerns about data quality, participants were excluded if fewer than 50% of trials were accepted in a particular task. This included three participants (one White woman, one White man, and one multiracial man), resulting in a final sample of 62 participants who contributed data to the analyses. Accuracy in the race-categorization task was high (*M* = 92.2%, *SD* = 6.2%).

Results

Mean RTs and P2 amplitudes separated by stimulus type are shown in Figure 2. Analyses of the means can be found in the supplementary material. First, we present analyses regarding

Table 1. Between- and Within-Person Effects of P2 Amplitude on Reaction Time for Categorization Decisions (Male/Female or Black/White) in Each Study.

	Study 1		Study 2				Study 3			
	Race task		Race task		Gender task		Race task		Gender task	
	n = 62		*n* = 63		*n* = 63		*n* = 65		*n* = 64	
	b	*p*	*b*	*p*	*b*	*p*	*b*	*P*	*b*	*p*
Intercept	457.4	<**.001**	435.3	<**.001**	457.6	<**.001**	437.5	<**.001**	450.3	<**.001**
Between-person effect	−0.03	.926	−0.01	.943	−0.02	.947	−0.03	.923	0.00	.989
Within-person effect	−0.44	<**.001**	−0.33	<**.001**	−0.32	<**.001**	−0.19	<**.001**	−0.18	<**.001**

Note. Bold items signify a significant *p* value (<.05).

the total effect of P2 amplitude on RT, partitioning between- and within-person effects, followed by analyses examining P3 latency as a mediator of this relationship.

Total effect of P2 amplitude on RT. As seen in Table 1, the within-person effect of P2 amplitude on RT was negative and significant, $b = -0.44$, $z = -14.01$, $p < .001$, such that a larger P2 amplitude on a particular trial (relative to a person's mean P2 amplitude) predicted a faster Black/White categorization response on that trial. The between-person effect was not significant ($p = .926$).

P3 latency as a mediator. When examining within-person effects, the a path (the effect of P2 amplitude on P3 latency) was significant, $b = -0.31$, $z = -12.17$, $p < .001$, as was the b path (the effect of P3 latency on RT, accounting for P2 amplitude), $b = 0.29$, $z = 73.5$, $p < .001$. The indirect effect was also significant, $ab = -0.088$ [−0.103, −0.074]. This relationship suggests that larger P2 amplitudes (relative to a person's average P2 amplitude) facilitate faster P3 latencies, and that faster P3 latencies (relative to a person's average P3 latency) facilitate faster RTs. However, the direct effect (the effect of P2 amplitude on RT, accounting for P3 latency) was still significant, $b = -0.35$, $z = -11.46$, $p < .001$, suggesting that P3 latency only partially mediates the within-person relationship between P2 amplitude and RT.

In contrast, when examining between-person effects, P2 amplitude did not predict P3 latency, $b = 0.02$, $z = 0.06$, $p = .954$, and P3 latency did not predict RT (controlling for P2 amplitude), $b = 0.03$, $z = 0.79$, $p = .433$. The between-person indirect effect was also not significant, $ab = 0.000$ [−0.022, 0.024].

Study 2

Participants

Sixty-six men participated in exchange for credit toward a course requirement or for monetary compensation. They ranged in age from 18 to 28 years ($M = 19.8$). Thirty-two participants self-identified as White and 34 self-identified as African American or Black. None self-identified as Hispanic. Three participants' data (two Black and one White) were excluded in the gender-categorization task and three participants' data (one Black and two White) were excluded in the race-categorization task for data quality concerns because more than 50% of trials were excluded. This left final samples of 63 participants each, contributing data to the analyses in each task. Accuracy was high in both the race-categorization task ($M = 91.8\%$, $SD = 5.3\%$) and the gender-categorization task ($M = 89.1\%$, $SD = 5.5\%$).

Results

Mean RTs and P2 amplitudes separated by stimulus type and participant race are shown in Figure 3. Detailed analyses regarding the means can be found in the supplementary material.

Total effect of P2 amplitude on RT. As in Study 1, the within-person effect of P2 amplitude on RT was large and significant both when categorizing faces by race, $b = -0.33$, $z = -13.16$, $p < .001$, and when categorizing faces by gender, $b = -0.32$, $z = -11.57$, $p < .001$ (see Table 1). The relationship was negative, such that a larger P2 amplitude to a face predicted a faster RT for both the Black/White categorization response and the male/female categorization response. The between-person effect was not significant in either task (ps = .943 and .947, respectively).

P3 latency as a mediator. Overall, patterns replicated the patterns seen in Study 1. When examining within-person effects in the race-categorization task, the a path (the effect of P2 amplitude on P3 latency) was significant, $b = -0.15$, $z = -6.82$, $p < .001$, as was the b path (the effect of P3 latency on RT, accounting for P2 amplitude), $b = 0.24$, $z = 80.40$, $p < .001$, and the indirect effect, $ab = -0.036$ [−0.047, −0.026]. The direct effect (the effect of P2 amplitude on RT, accounting for P3 latency) was still significant, $b = -0.29$, $z = -11.74$, $p < .001$, suggesting that P3 latency is only a partial mediator of the within-person relationship between P2 amplitude and RT in the race-categorization task. In the gender-categorization task, as in the race-categorization task, both the a path, $b = -0.12$, $z = -5.45$, $p < .001$, and the

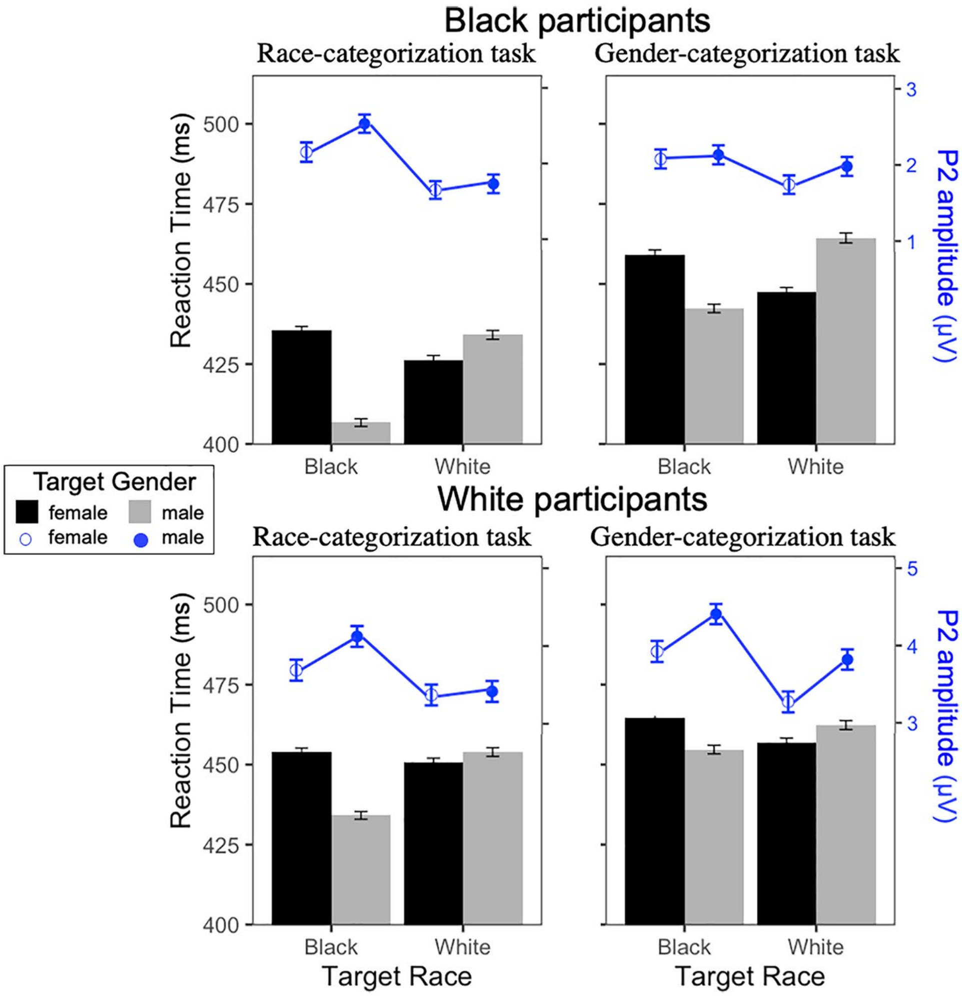

Figure 3. Mean RTs and mean P2 amplitude by stimulus condition in Study 2.

Note. Black and gray bars correspond to the *y*-axis on the left (reaction time) and blue points correspond to the *y*-axis on the right (P2 amplitude). Error bars represent 95% CIs. $N = 63$. Participants were White (race-categorization task: $n = 30$; gender-categorization task: $n = 31$) or Black (race-categorization task: $n = 33$; gender-categorization task: $n = 32$). All participants were men. RT = reaction time; CI = confidence interval.

b path, $b = 0.26$, $z = 77.85$, $p < .001$, were significant, as well as the indirect effect, $ab = -0.032$ [−0.043, −0.020]. The direct effect was still significant, $b = -0.29$, $z = -10.62$, $p < .001$.

When examining between-person effects, results differed slightly from Study 1. Unlike Study 1, P2 amplitude significantly predicted P3 latency in both the race-categorization task, $b = -1.35$, $z = -7.58$, $p < .001$, and the gender-categorization task, $b = -1.38$, $z = -7.80$, $p < .001$. However, P3 latency did not significantly predict RT (controlling for P2 amplitude) in either the race-categorization task, $b = 0.03$, $z = 0.92$, $p = .357$, or the gender-categorization task, $b = 0.04$, $z = 1.22$, $p = .224$. Thus, as in Study 1, the between-person indirect effect was not significant in either task (race: −0.038 [−0.122, 0.043]; gender: −0.053 [−0.143, 0.032]).

Study 3

Participants

Sixty-eight individuals (43 women, 21 men, three transgender/nonbinary people, and one did not report) completed the race- and gender-categorization tasks while EEG was recorded as a portion of a larger study. They received monetary compensation for their participation and ranged in age from 18 to 45 years ($M = 20.6$). All self-identified as African American or Black. One participant (male) did not complete the gender-categorization task because of technical malfunction. Three additional participants (all women) were excluded from analysis due to low data quality (>50% of trials excluded). As a result, 65 participants contributed data to analyses of the race-categorization task and 64 participants contributed data to analyses of the gender-categorization

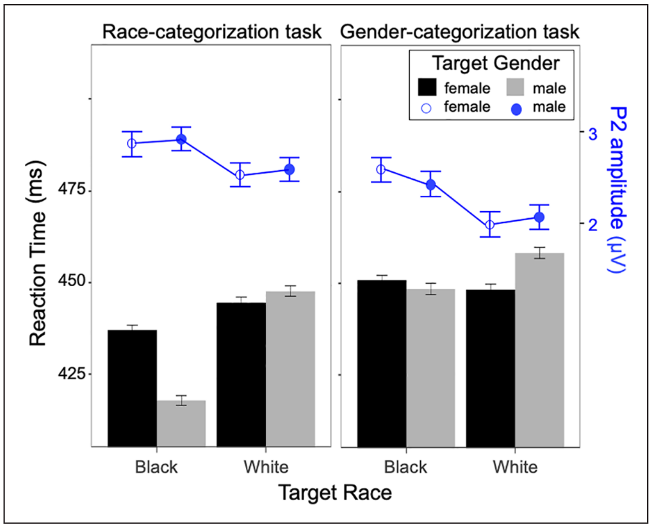

Figure 4. Mean RTs and mean P2 amplitude by stimulus condition in Study 3.

Note. White and gray bars correspond with *y*-axis on the left (reaction time) and blue points correspond with *y*-axis on the right (P2 amplitude). Error bars represent 95% CIs. All participants were Black (N = 65 in race-categorization task; N = 64 in gender-categorization task); women comprised the majority of the sample. RT = reaction time; CI = confidence interval.

task. Accuracy was high in both the race-categorization task (M = 92.8%, SD = 6.1%) and the gender-categorization task (M = 90.6%, SD = 6.4%).

Results

Mean RTs separated by stimulus type are shown in Figure 4. Detailed analyses regarding the means can be found in the supplementary material.

Total effect of P2 amplitude on RT. As in Studies 1 and 2, the within-person effect of P2 amplitude on RT was significant in both the race-categorization task, $b = -0.19$, $z = -4.73$, $p < .001$, and the gender-categorization task, $b = -0.18$, $z = -4.12$, $p < .001$ (see Table 1). The larger a P2 amplitude elicited by a particular face (relative to a person's average P2 amplitude), the faster the RT when making a Black/White categorization or a male/female categorization of that face. There was no between-person effect in either task (ps = .923 and .989, respectively).

P3 latency as a mediator. As in Studies 1 and 2, when examining within-person effects in the race-categorization task, the a path was significant, $b = -0.25$, $z = -7.26$, $p < .001$, as was the b path, $b = 0.26$, $z = 54.59$, $p < .001$, and the indirect effect, $ab = -0.063$ [−0.080, −0.046]. As in Studies 1 and 2, the direct effect was still significant, $b = -0.13$, $z = -3.24$, $p = .001$, suggesting partial mediation. In the gender-categorization task, the same pattern emerged. Both the a path, $b = -0.11$, $z = -3.26$, $p = .001$, and the b path, $b = 0.25$, $z = 50.84$, $p < .001$, were significant, as well as both the indirect effect, $ab = -0.029$ [−0.046, −0.011], and the direct effect, $b = -0.15$, $z = -3.55$, $p < .001$.

When examining between-person effects, P2 amplitude predicted P3 latency in both race-categorization task, $b = -1.41$, $z = -6.15$, $p < .001$, and the gender-categorization task, $b = -1.66$, $z = -7.22$, $p < .001$, as in Study 2. In addition, P3 latency significantly predicted RT (controlling for P2 amplitude) in the race-categorization task, $b = 0.07$, $z = 2.34$, $p = .020$, but not the gender-categorization task, $b = 0.06$, $z = 1.38$, $p = .169$. The between-person indirect effect was significant in the race-categorization task, $ab = -0.105$ [−0.206, −0.016] but not the gender-categorization task, $ab = -0.091$ [−0.229, 0.038].

Discussion

The manner in which people derive information about others' social category memberships from their faces has long been of interest to social and cognitive psychologists (Bodenhausen et al., 2012; Cloutier et al., 2005; Macrae & Bodenhausen, 2000). These categorization decisions have profound implications for downstream evaluations, activation of stereotypes, and motivation of interpersonal responses (Eberhardt et al., 2006; Kawakami et al., 2017; Quinn & Rosenthal, 2012; Zebrowitz, 2006). Both the DI model of person construal and CF models of information processing suggest that information about a stimulus accumulates over time, and that the amount of task-relevant information extracted early on should determine the efficiency of stimulus evaluation (Coles et al., 1985; Freeman & Ambady, 2011; Freeman & Johnson, 2016; Gratton et al., 1988). In the current studies, we provide novel evidence using ERPs of the relationship between early extraction of category-relevant information and subsequent facilitation of goal-relevant stimulus evaluation under two task goals (i.e., the categorization of faces by race and by gender).

Specifically, we reasoned that the face-elicited P2, long known to be associated with early allocation of selective attention (Hillyard & Münte, 1984; Luck & Hillyard, 1994; Mangun, 1995; Wijers et al., 1989) and highly sensitive to social category distinctions (e.g., Correll et al., 2006; He et al., 2009; Ito & Urland, 2003, 2005; Volpert-Esmond & Bartholow, 2019), should reflect early extraction of goal-relevant information during person construal, thereby facilitating overt assignment of a face to a relevant category. To date, evidence in support of this functional relationship has been elusive (Dickter & Bartholow, 2007; Kubota & Ito, 2007), which we argue reflects the reliance on signal averaging and tests of between-person associations in previous studies. By testing for associations at the level of individual trials and disaggregating between- and within-person sources of variability (Curran & Bauer, 2011), the current studies provided the first test of whether within-person variability in P2 amplitude predicts the speed with which social categories are evaluated (P3 latency) and overt responses are made (RT).

Consistent with previous work (Dickter & Bartholow, 2007; Kubota & Ito, 2007), no between-person effect of P2 amplitude on RT was observed. In other words, having a larger-than-average P2 amplitude, relative to other people, did not facilitate stimulus evaluation or categorization responses. This may reflect the contribution of nonpsychological physiological factors, such as skull thickness (Chauveau et al., 2004; Hagemann et al., 2008), to between-person variance, increasing the power needed to detect a functional between-person relationship.

Critically, despite differences across studies in the samples (Study 1: roughly half men and women, mostly White; Study 2: all men, roughly 50% Black; Study 3: roughly two-thirds women, all Black), stimuli (some sets of only male faces; some sets of male and female faces), and experimental manipulations (varying top-down categorization goals), within-person, trial-to-trial variation in face-elicited P2 amplitude meaningfully predicted subsequent categorization RTs in all three studies. In other words, a larger-than-average P2 amplitude on a particular trial, relative to an individual's mean P2 amplitude, facilitated a faster categorization judgment on that trial. Interestingly, this within-person relationship was evident regardless of whether participants categorized faces by race or by gender, suggesting that P2 amplitude is sensitive to task-relevant goals and facilitates the categorization response across multiple types of categorization. This is consistent with previous research (e.g., Volpert-Esmond & Bartholow, 2019) showing that the P2 is subject to top-down task goals and thus may reflect early extraction of whatever information is goal-relevant.

Moreover, the within-person relationship between P2 amplitude and categorization RT was mediated by the latency of the P3. We hypothesized that extraction of more category-relevant information early in the processing stream (i.e., larger P2 amplitude) would facilitate the accumulation of information, resulting in more efficient stimulus evaluation (i.e., P3 latency), which then would facilitate faster generation of overt categorization responses. Consistent with this prediction, P3 latency significantly mediated the within-person relationship between P2 amplitude and RT across studies and tasks. This finding further supports the hypothesis that extraction of more category-relevant information early in processing facilitates the evaluation of a face into task-relevant social categories on a given trial, thereby enhancing person construal.

Thus far, consistent with numerous other face-perception studies (for a review, see Ito & Bartholow, 2009), we have referred to this early component as the P2, conceptualizing it as very similar to the attention-related P2 component described in studies investigating early allocation of selective attention (Hillyard & Münte, 1984; Luck & Hillyard, 1994; Mangun, 1995; Wijers et al., 1989). However, there are similarities between this P2 and the vertex positive potential (VPP), which is the inversion of the face-elicited N170 ERP component (Joyce & Rossion, 2005). The N170/VPP is larger to face stimuli compared with non-face stimuli and has been proposed to index structural encoding of faces (Eimer, 2000). Given that the latency of the early component we report in these studies (measured 130–190 ms post-face onset) is similar both in latency to the N170/VPP (generally within 140–180 ms post-face onset) and scalp location (fronto-centro-parietal sites), it is reasonable to wonder whether the component traditionally called the face-elicited P2 is in fact the N170/VPP.

However, we believe that the face-elicited P2 described in our studies is distinct from the N170/VPP for two primary reasons. First, whereas the face-elicited P2 is consistently larger to racial outgroup faces compared with racial ingroup faces across dozens of studies, regardless of task (e.g., Ito & Tomelleri, 2017; for a review see Ito & Bartholow, 2009), the literature is more mixed regarding the effect of racial group membership on the N170/VPP. Some studies report no difference in N170/VPP amplitude based on racial group membership (e.g., Caldara et al., 2003, 2004; Chen et al., 2013; Ito et al., 2004; Ofan et al., 2011), others report larger N170/VPP amplitude to outgroup faces compared with ingroup faces (He et al., 2009; Herrmann et al., 2007; Stahl et al., 2008, 2010; Volpert-Esmond et al., 2018; Walker et al., 2008; Wiese et al., 2014), and still others report opposing patterns of N170/VPP amplitude to ingroup/outgroup faces depending on task parameters (e.g., categorization vs. identification, Senholzi & Ito, 2013). In addition, one study examined the N170 and P2 concurrently and showed separate patterns in the effect of skin color and facial physiognomy on the two components (Brebner et al., 2011). Thus, the face-elicited P2 and the N170/VPP appear to be somewhat functionally different with regard to social categorization.

Second, whereas principal component analysis (PCA) has shown that the face-elicited P2 reflects two functionally separate underlying components (Volpert-Esmond & Bartholow, 2019; Volpert-Esmond et al., 2017), research applying PCA to the N170/VPP has shown that it reflects a single underlying component (Ito et al., 2004), reflecting activity in a single neural generator (Deffke et al., 2007; Rossion et al., 2003). For these reasons, we believe that the face-elicited P2 is functionally distinct from the N170/VPP. It would be valuable for future research to specify similarities and differences between the face-elicited P2 and the N170/VPP to determine whether the P2 elicited in social categorization studies and the VPP typically observed in face processing studies are, in fact, the same component and reflect similar underlying neurophysiological activity.

Thus, we assume that the P2 reflects variation in the extent to which attention is allocated, facilitating extraction of category-relevant information early in the processing stream, which then affects the duration of stimulus evaluation and, ultimately, the speed of overt categorization. A naturally following question is *why* attention and early extraction of category-relevant information (i.e., P2 amplitude) varied across trials. One possibility is that differences in conditions, such as where a perceiver fixates on a given face, allow for more or less extraction of category-relevant information from the face. Studies suggest that different facial features convey category-relevant information to different degrees (Hills & Lewis, 2006, 2011), and previous studies have shown both an increase in P2 amplitude and speeding of categorization RT when participants fixate between the eyes compared with on the forehead of a face (Volpert-Esmond et al., 2017, 2019; more information regarding the effect of fixation in the current studies can be found in the supplementary material). Thus, future research should examine the possibility that the availability of social category information in the features attended to influences variation in P2 amplitude across trials.

In addition, properties of the face stimuli themselves may contribute to variation in P2 amplitude. One interpretation of the well-known pattern of larger P2 amplitudes to Black faces than White faces among White participants is that Black faces are more salient to White participants (e.g., Ito & Bartholow, 2009). Additional research has shown mean differences in P2 amplitude as function of emotional expression, such that both angry and happy faces elicited larger P2 amplitudes (Kubota & Ito, 2007). These studies suggest that P2 amplitude may differ as a function of the salience of a particular face, which then facilitates extraction of social category–relevant information. Other properties of the faces themselves may additionally contribute to variation in P2 amplitude, including how prototypical a face is, or other variability in facial features. Future research should examine how qualities present in the faces contribute to early attention and resulting category decisions.

Finally, our analyses suggest that a given participant's P2 responses to the same face—let alone to different faces—vary across trials, and that this variability is functionally important. A likely contributor to this variability is that task engagement and control of attention simply vary over the course of the experiment. It has long been known that intertrial variation in RT is attributable to fluctuations in sustained attention (Flehmig et al., 2007; Jensen, 1992; Sanders, 1998). Research has identified the neural bases for these fluctuations in both pre-stimulus decreases in prefrontal regions involved in controlling attention and in reduced stimulus-elicited sensory activation (Weissman et al., 2006). Given the P2's sensitivity to top-down task demands and perceiver characteristics (Correll et al., 2006; He et al., 2009; Volpert-Esmond & Bartholow, 2019), future research could manipulate factors known to influence sustained attention and test their influence on the P2 and subsequent stimulus evaluation and response output measures.

One unexpected finding is in regard to the pattern in RT to racial ingroup and outgroup faces. Previous literature has reported an "other-race advantage" in classifying faces by race (Caldara et al., 2004; Feng et al., 2011; Levin, 1996). A number of theoretical accounts have been proposed, including that recognition or individuation processes conflict with categorization processes, resulting in enhanced individuation (e.g., Meissner & Brigham, 2001) and detriments to categorization (Levin, 1996; Valentine & Endo, 1992) for racial ingroup faces. However, across all three studies, we did not observe a consistent response time facilitation for racial outgroup relative to ingroup faces. White participants showed the typical other-race advantage in categorization speed when categorizing male faces in Studies 1 and 2, but not female faces in Study 2. Black participants showed the typical other-race advantage only when categorizing female faces in Study 2 but not in any other case (see the supplementary material, for detailed analyses regarding RT). Because of the novelty of these results, it is unclear why we do not see a consistent pattern.

One possibility is that the current studies differ with regard to the context and familiarity that participants have with the racial outgroup. Previous studies demonstrating the other-race advantage in categorization speed have primarily relied on participants with little to no experience or contact with individuals from the racial outgroup (Caharel et al., 2011; Caldara et al., 2004; Feng et al., 2011; Ge et al., 2009; Zhao & Bentin, 2008). In the current research, Black participants attended a predominantly White institution in the United States where they presumably had extensive experience seeing and interacting with White individuals. White participants also likely had at least some experience seeing and interacting with Black individuals although previous experience with the racial outgroup was not measured in these studies. It is becoming increasingly clear that the interplay between categorization and individuation likely depends on sociocultural factors (Hugenberg et al., 2010) and that racial categorization depends on characteristics of the perceiver (Kawakami et al., 2017). Thus, the racial dynamics the participants are embedded in and the experience they have with racial outgroup members may contribute to the pattern of results seen in this study. Future research should investigate how racial environments and previous experience with racial outgroups impact how quickly participants classify racial ingroup and outgroup faces, and investigate the possibility that the other-race advantage in categorization is not as universal as previously thought.

In conclusion, this is the first demonstration of the face-elicited P2's functional significance for social categorization. Although we focus on the social categorization of faces, we believe that similar effects may be found when categorizing other types of objects into meaningful groups, as is evident in other studies testing the CF model. This demonstration was made possible by the novel use of single-trial ERPs and multilevel models that allowed for disaggregation of within- and between-person effects in the relationship between P2 amplitude, stimulus evaluation, and categorization RTs. Mapping out functional relationships between brain and behavior has the potential to contribute to understanding of the real-life consequences of categorization, including when police officers are influenced by race when making split-second decisions to shoot (Correll et al., 2002, 2006, 2007) or when other evaluative or stereotypic biases related to social category affect interpersonal interactions. Future research may examine how top-down perceiver goals or contextual influences moderate the relationship between brain and behavior.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported, in part, by a Life Sciences Graduate Fellowship from the University of Missouri, a National Research Service Award from the National Institute on Minority Health and Health Disparities (F31 MD012751), and a Research Council grant from the University of Missouri. BDB's contribution to the preparation of the manuscript was additionally supported by the National Institute on Alcohol Abuse and Alcoholism (R01 AA025451).

ORCID iD

Hannah I. Volpert-Esmond https://orcid.org/0000-0002-9181-9710

Supplemental Material

Supplemental material is available online with this article.

Notes

1. More information about the fixation manipulation can be found in Volpert-Esmond et al. (2017) and Volpert-Esmond and Bartholow (2019).
2. Study 1: FP1, FP2, Fz, FCz, FC3, FC4, Cz, C3, C4, CPz, CP3, CP4, Pz, P3, P4, Oz, TP7, TP8, P7/T5, and P8/T6. Study 2: FP1, FP2, Fz, F1, F2, F3, F4, FCz, FC3, FC4, Cz, C1, C2, C3, C4, CPz, CP1, CP2, CP3, CP4, Pz, P3, P4, POz, PO3, PO4, PO5, PO6, Oz, O1, O2, TP7, TP8, T5/P7, and T6/P8. Study 3: FP1, FP2, Fz, F1, F2, F3, F4, FCz, FC3, FC4, Cz, C1, C2, C3, C4, CPz, CP3, CP4, Pz, P1, P2, P3, P4, POz, PO5, PO6, PO7, PO8, Oz, TP7, TP8, T5/P7, and T6/P8. Additional electrodes were placed on the tip of the nose (Study 2 only), above and below the left eye and on the outer canthus of each eye (to record blinks and saccades) and over each mastoid.
3. This was reported incorrectly in Volpert-Esmond et al. (2017) as 10 to 40 Hz.
4. For Study 2, this was reported incorrectly in Volpert-Esmond and Bartholow (2019) as .10 to .40 Hz.
5. Wilkinson notation: RT ~ P2.within + P2.between + (1|Subject).
6. Wilkinson notation: P3 ~ P2.within + P2.between + (1|Subject).
7. Wilkinson notation: RT ~ P2.within + P2.between + P3.within + P3.between + (1|Subject).

References

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*(2), 231–239. https://doi.org/10.1016/S0959-4388(00)00202-6

Amodio, D. M. (2009). Coordinated roles of motivation and perception in the regulation of intergroup responses: Frontal cortical asymmetry effects on the P2 event-related potential and behavior. *Journal of Cognitive Neuroscience*, *22*(11), 2609–2617. https://doi.org/10.1162/jocn.2009.21395

Amodio, D. M., Bartholow, B. D., & Ito, T. A. (2014). Tracking the dynamics of the social brain: ERP approaches for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, *9*(3), 385–393. https://doi.org/10.1093/scan/nst177

Bargh, J. A. (2014). The historical origins of priming as the preparation of behavioral responses: Unconscious carryover and contextual influences of real-world importance. *Social Cognition*, *32*(Supplement), 209–224. https://doi.org/10.1521/soco.2014.32.supp.209

Bartholow, B. D., Riordan, M. A., Saults, J. S., & Lust, S. A. (2009). Psychophysiological evidence of response conflict and strategic control of responses in affective priming. *Journal of Experimental Social Psychology*, *45*(4), 655–666. https://doi.org/10.1016/j.jesp.2009.02.015

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berry, M. P., Tanovic, E., Joormann, J., & Sanislow, C. A. (2019). Relation of depression symptoms to sustained reward and loss sensitivity. *Psychophysiology*, *56*(7), Article e13364. https://doi.org/10.1111/psyp.13364

Bodenhausen, G. V., Kang, S. K., & Peery, D. (2012). Social categorization and the perception of social groups. In S. T. Fiske & C. N. Macrae (Eds.), *The Sage handbook of social cognition* (pp. 318–336). SAGE.

Brebner, J. L., Krigolson, O., Handy, T. C., Quadflieg, S., & Turk, D. J. (2011). The importance of skin color and facial structure in perceiving and remembering others: An electrophysiological study. *Brain Research*, *1388*, 123–133. https://doi.org/10.1016/j.brainres.2011.02.090

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123–152. https://doi.org/10.1037/h0043805

Brush, C. J., Ehmann, P. J., Hajcak, G., Selby, E. A., & Alderman, B. L. (2018). Using multilevel modeling to examine blunted neural responses to reward in major depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(12), 1032–1039. https://doi.org/10.1016/j.bpsc.2018.04.003

Caharel, S., Montalan, B., Fromager, E., Bernard, C., Lalonde, R., & Mohamed, R. (2011). Other-race and inversion effects during the structural encoding stage of face processing in a race categorization task: An event-related brain potential study. *International Journal of Psychophysiology*, *79*(2), 266–271. https://doi.org/10.1016/j.ijpsycho.2010.10.018

Caldara, R., Rossion, B., Bovet, P., & Hauert, C. A. (2004). Event-related potentials and time course of the "other-race" face classification advantage. *NeuroReport*, *15*(5), 905–910.

Caldara, R., Thut, G., Servoir, P., Michel, C. M., Bovet, P., & Renault, B. (2003). Face versus non-face object perception and the "other-race" effect: A spatio-temporal event-related potential study. *Clinical Neurophysiology*, *114*(3), 515–528. https://doi.org/10.1016/S1388-2457(02)00407-8

Chauveau, N., Franceries, X., Doyon, B., Rigaud, B., Morucci, J. P., & Celsis, P. (2004). Effects of skull thickness, anisotropy, and inhomogeneity on forward EEG/ERP computations using a spherical three-dimensional resistor mesh model. *Human Brain Mapping*, *21*(2), 86–97. https://doi.org/10.1002/hbm.10152

Chen, Y., Pan, F., Wang, H., Xiao, S., & Zhao, L. (2013). Electrophysiological correlates of processing own- and other-race faces. *Brain Topography*, *26*(4), 606–615. https://doi.org/10.1007/s10548-013-0286-x

Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology*, *88*(6), 885–894. https://doi.org/10.1037/0022-3514.88.6.885

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology. Human Perception and Performance*, *11*(5), 529–553. https://doi.org/10.1037//0096-1523.11.5.529

Coles, M. G. H., Gratton, G., & Fabiani, M. (1990). Event-related brain potentials. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements* (pp. 413–455). Cambridge University Press.

Coles, M. G. H., & Rugg, M. D. (1995). Event-related potentials: An introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 1–26). Oxford University Press.

Coles, M. G. H., Smid, H. G. O. M., Scheffers, M. K., & Otten, L. J. (1995). Mental chronometry and the study of human information processing. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 86–131). Oxford University Press.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314–1329. https://doi.org/10.1037/0022-3514.83.6.1314

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, *92*(6), 1006–1023. https://doi.org/10.1037/0022-3514.92.6.1006

Correll, J., Urland, G. R., & Ito, T. A. (2006). Event-related potentials and the decision to shoot: The role of threat perception and cognitive control. *Journal of Experimental Social Psychology*, *42*(1), 120–128. https://doi.org/10.1016/j.jesp.2005.02.006

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*(1), 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Dawson, G. D. (1954). A summation technique for the detection of small evoked potentials. *Electroencephalography & Clinical Neurophysiology*, *6*, 65–84. https://doi.org/10.1016/0013-4694(54)90007-3

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience*, *25*(50), 11730–11737. https://doi.org/10.1523/JNEUROSCI.3286-05.2005

Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L., & Lueschow, A. (2007). MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. *NeuroImage*, *35*(4), 1495–1501. https://doi.org/10.1016/j.neuroimage.2007.01.034

Dickter, C. L., & Bartholow, B. D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience*, *2*(3), 189–198. https://doi.org/10.1093/scan/nsm012

Dickter, C. L., & Kittel, J. A. (2012). The effect of stereotypical primes on the neural processing of racially ambiguous faces. *Social Neuroscience*, *7*(6), 622–631. https://doi.org/10.1080/17470919.2012.690345

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked potentials and behavior* (pp. 13–75). Plenum Press.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–374. https://doi.org/10.1017/S0140525X00058027

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, *14*(1), 1–52. https://doi.org/10.1016/0301-0511(82)90016-3

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383–386. https://doi.org/10.1111/j.1467-9280.2006.01716.x

Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces. *NeuroReport*, *11*(10), 2319–2324.

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, *25*(4), 249–263. https://doi.org/10.3758/BF03198804

Feng, L., Liu, J., Wang, Z., Li, J., Li, L., Ge, L., Tian, J., & Lee, K. (2011). The other face of the other-race effect: An fMRI investigation of the other-race face categorization advantage. *Neuropsychologia*, *49*(13), 3739–3749. https://doi.org/10.1016/j.neuropsychologia.2011.09.031

Flehmig, H. C., Steinborn, M., Langner, R., Scholz, A., & Westhoff, K. (2007). Assessing intraindividual variability in sustained attention: Reliability, relation to speed and accuracy, and practice effects. *Psychology Science*, *49*(2), 132–149.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279.

Freeman, J. B., & Ambady, N. (2014). The dynamic interactive model of person construal: Coordinating sensory and social processes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 235–248). The Guilford Press.

Freeman, J. B., Ambady, N., & Holcomb, P. J. (2010). The face-sensitive N170 encodes social category information. *NeuroReport*, *21*(1), 24–28. https://doi.org/10.1097/WNR.0b013e3283320d54

Freeman, J. B., Ambady, N., Midgley, K. J., & Holcomb, P. J. (2011). The real-time link between person perception and action: Brain potential evidence for dynamic continuity. *Social Neuroscience*, *6*(2), 139–155. https://doi.org/10.1080/17470919.2010.490674

Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences*, *20*(5), 362–374. https://doi.org/10.1016/j.tics.2016.03.003

Freeman, J. B., Pauker, K., & Sanchez, D. T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science*, *27*(4), 502–517. https://doi.org/10.1177/0956797615627418

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape racial perception. *PLOS ONE*, *6*(9), Article e25107.

Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., Slater, A., Tian, J., & Lee, K. (2009). Two faces of the other-race effect: Recognition and categorisation of Caucasian and Chinese faces. *Perception*, *38*(8), 1199–1210. https://doi.org/10.1068/p6136

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gratton, G., Bosco, C. M., Kramer, A. F., Coles, M. G. H., Wickens, C. D., & Donchin, E. (1990). Event-related brain potentials as indices of information extraction and response priming. *Electroencephalography and Clinical Neurophysiology*, *75*(5), 419–432. https://doi.org/10.1016/0013-4694(90)90087-Z

Gratton, G., Coles, M. G. H., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 331–344. https://doi.org/10.1037/0096-1523.14.3.331

Hackley, S. A. (2009). The speeding of voluntary reaction by a warning signal. *Psychophysiology*, *46*(2), 225–233. https://doi.org/10.1111/j.1469-8986.2008.00716.x

Hagemann, D., Hewig, J., Walter, C., & Naumann, E. (2008). Skull thickness and magnitude of EEG alpha activity. *Clinical Neurophysiology*, *119*(6), 1271–1280. https://doi.org/10.1016/j.clinph.2008.02.010

He, Y., Johnson, M. K., Dovidio, J. F., & McCarthy, G. (2009). The relation between race-related implicit associations and scalp-recorded neural activity evoked by faces from different races. *Social Neuroscience*, *4*(5), 426–442. https://doi.org/10.1080/17470910902949184

Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early processing of gendered facial cues predicts the electoral success of female politicians. *Social Psychological and Personality Science*, *5*(7), 815–824. https://doi.org/10.1177/1948550614534701

Herrmann, M. J., Schreppel, T., Jäger, D., Koehler, S., Ehlis, A.-C., & Fallgatter, A. J. (2007). The other-race effect for face perception: An event-related potential study. *Journal of Neural Transmission*, *114*(7), 951–957. https://doi.org/10.1007/s00702-007-0624-9

Hills, P. J., & Lewis, M. B. (2006). Reducing the own-race bias in face recognition by shifting attention. *The Quarterly Journal of Experimental Psychology*, *59*(6), 996–1002. https://doi.org/10.1080/17470210600654750

Hills, P. J., & Lewis, M. B. (2011). Reducing the own-race bias in face recognition by attentional shift using fixation crosses preceding the lower half of a face. *Visual Cognition*, *19*(3), 313–339. https://doi.org/10.1080/13506285.2010.528250

Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception and Pscyhophysics*, *36*, 185–198.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review*, *117*(4), 1168–1187. https://doi.org/10.1037/a0020463

Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in Cognitive Sciences*, *13*(12), 524–531. https://doi.org/10.1016/j.tics.2009.10.002

Ito, T. A., Thompson, E., & Cacioppo, J. T. (2004). Tracking the timecourse of social perception: The effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin*, *30*(10), 1267–1280. https://doi.org/10.1177/0146167204264335

Ito, T. A., & Tomelleri, S. (2017). Seeing is not stereotyping: The functional independence of categorization and stereotype activation. *Social Cognitive and Affective Neuroscience*, *12*(5), 758–764. https://doi.org/10.1093/scan/nsx009

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*(4), 616–626. https://doi.org/10.1037/0022-3514.85.4.616

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 21–36. https://doi.org/10.3758/CABN.5.1.21

Jensen, A. R. (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences*, *13*(8), 869–881. https://doi.org/10.1016/0191-8869(92)90004-9

Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. *Clinical Neurophysiology*, *116*(11), 2613–2631. https://doi.org/10.1016/j.clinph.2005.07.005

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. *Advances in Experimental Social Psychology*, *55*, 1–80. https://doi.org/10.1016/bs.aesp.2016.10.001

Kramer, R. S. S., Young, A. W., Day, M. G., & Burton, A. M. (2017). Robust social categorization emerges from learning the identities of very few faces. *Psychological Review*, *124*(2), 115–129. https://doi.org/10.1037/rev0000048

Kubota, J. T., & Ito, T. A. (2007). Multiple cues in social perception: The time course of processing race and facial expression. *Journal of Experimental Social Psychology*, *43*(5), 738–752. https://doi.org/10.1016/j.jesp.2006.10.023

Kubota, J. T., & Ito, T. A. (2017). Rapid race perception despite individuation and accuracy goals. *Social Neuroscience*, *12*(4), 468–478. https://doi.org/10.1080/17470919.2016.1182585

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792–795. https://doi.org/10.1126/science.887923

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1364–1382. https://doi.org/10.1037/0278-7393.22.6.1364

Levsen, M. P., Volpert-Esmond, H. I., & Bartholow, B. D. (2018). It's about time: Event-related potentials and the temporal parameters of mental events. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in social psychology* (pp. 102–126). Taylor & Francis.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, *66*(1), 735–769. https://doi.org/10.1146/annurev.physiol.66.082602.092845

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(5), 1000–1014. https://doi.org/10.1037//0096-1523.20.5.1000

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*(1), 93–120. https://doi.org/10.1146/annurev.psych.51.1.93

Macrae, C. N., Quinn, K. A., Mason, M. F., & Quadflieg, S. (2005). Understanding others: The face and person construal. *Journal of Personality and Social Psychology*, *89*(5), 686–695. https://doi.org/10.1037/0022-3514.89.5.686

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 4–18. https://doi.org/10.1111/j.1469-8986.1995.tb03400.x

Mason, M. F., & Macrae, C. N. (2004). Categorizing and individuating others: The neural substrates of person perception. *Journal of Cognitive Neuroscience*, *16*(10), 1785–1795. https://doi.org/10.1162/0898929042947801

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, *211*(4477), 77–80. https://doi.org/10.1126/science.7444452

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, *23*(1), 1–44. https://doi.org/10.1016/0010-0285(91)90002-6

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus—Norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. https://doi.org/10.1037/0033-2909.131.4.510

Ofan, R. H., Rubin, N., & Amodio, D. M. (2011). Seeing race: N170 responses to race and their relation to automatic racial attitudes and controlled processing. *Journal of Cognitive Neuroscience*, *23*(10), 3153–3161. https://doi.org/10.1162/jocn_a_00014

Pernet, C. R., Sajda, P., & Rousselet, G. A. (2011). Single-trial analyses: Why bother? *Frontiers in Psychology*, *2*, Article 322. https://doi.org/10.3389/fpsyg.2011.00322

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience*, *26*(35), 8965–8975. https://doi.org/10.1523/JNEUROSCI.1655-06.2006

Pituch, K. A., & Stapleton, L. M. (2008). The performance of methods to test upper-level mediation in the presence of nonnormal data. *Multivariate Behavioral Research*, *43*(2), 237–267. https://doi.org/10.1080/00273170802034844

Pituch, K. A., Whittaker, T. A., & Stapleton, L. M. (2005). A comparison of methods to test for mediation in multisite experiments. *Multivariate Behavioral Research*, *40*(1), 1–23. https://doi.org/10.1207/s15327906mbr4001_1

Polich, J. (2012). Neuropsychology of P300. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford handbook of event-related potential components* (pp. 159–188). Oxford University Press.

Polich, J., Ellerson, P. C., & Cohen, J. (1996). P300, stimulus intensity, modality, and probability. *International Journal of Psychophysiology*, *23*(1), 55–62. https://doi.org/10.1016/0167-8760(96)00028-1

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology*, *66*(1), 825–852. https://doi.org/10.1146/annurev-psych-010814-015258

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. https://doi.org/10.1037/a0020141

Quinn, K. A., & Rosenthal, H. E. S. (2012). Categorizing others and the self: How social memory structures guide social perception and behavior. *Learning and Motivation*, *43*(4), 247–258. https://doi.org/10.1016/j.lmot.2012.05.008

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences*, *106*(16), 6539–6544. https://doi.org/10.1073/pnas.0812589106

Regtvoort, A. G. F. M., van Leeuwen, T. H., Stoel, R. D., & van der Leij, A. (2006). Efficiency of visual information processing in children at-risk for dyslexia: Habituation of single-trial ERPs. *Brain and Language*, *98*(3), 319–331. https://doi.org/10.1016/j.bandl.2006.06.006

Requin, J., Riehle, A., & Seal, J. (1988). Neuronal activity and information processing in motor control: From stages to continuous flow. *Biological Psychology*, *26*(1), 179–198. https://doi.org/10.1016/0301-0511(88)90019-1

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage*, *20*(3), 1609–1624. https://doi.org/10.1016/j.neuroimage.2003.07.010

Rule, N. O., & Ambady, N. (2010). Democrats and republicans can be differentiated from their faces. *PLOS ONE*, *5*(1), Article e8733. https://doi.org/10.1371/journal.pone.0008733

Sanders, A. F. (1998). *Elements of human performance*. Lawrence Erlbaum.

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. https://doi.org/10.1016/j.bandl.2014.07.010

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, *23*(6), 695–703. https://doi.org/10.1111/j.1469-8986.1986.tb00696.x

Senholzi, K. B., & Ito, T. A. (2013). Structural face encoding: How task affects the N170's sensitivity to race. *Social Cognitive and Affective Neuroscience*, *8*(8), 937–942. https://doi.org/10.1093/scan/nss091

Simon, J. C., & Gutsell, J. N. (2019). Effects of minimal grouping on implicit prejudice, infrahumanization, and neural processing despite orthogonal social categorizations. *Group Processes & Intergroup Relations*, *23*, 323–343. https://doi.org/10.1177/1368430219837348

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *13*(2), 238–241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x

Stahl, J., Wiese, H., & Schweinberger, S. R. (2008). Expertise and own-race bias in face processing: An event-related potential study. *NeuroReport*, *19*(5), 583–587. https://doi.org/10.1097/WNR.0b013e3282f97b4d

Stahl, J., Wiese, H., & Schweinberger, S. R. (2010). Learning task affects ERP-correlates of the own-race bias, but not recognition memory performance. *Neuropsychologia*, *48*(7), 2027–2040. https://doi.org/10.1016/j.neuropsychologia.2010.03.024

Stolier, R. M., & Freeman, J. B. (2017). A neural mechanism of social categorization. *The Journal of Neuroscience*, *37*(23), 5711–5721. https://doi.org/10.1523/JNEUROSCI.3334-16.2017

Tibon, R., & Levy, D. A. (2015). Striking a balance: Analyzing unbalanced event-related potential data. *Frontiers in Psychology*, *6*, Article 555 https://doi.org/10.3389/fpsyg.2015.00555

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, *43*(3), 692–700. https://doi.org/10.3758/s13428-011-0076-x

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology Section A*, *44*(4), 671–703. https://doi.org/10.1080/14640749208401305

Volpert-Esmond, H. I., & Bartholow, B. D. (2019). Explicit categorization goals affect attention-related processing of race and gender during person construal. *Journal of Experimental Social Psychology*, *85*, Article 103839. https://doi.org/10.1016/j.jesp.2019.103839

Volpert-Esmond, H. I., Merkle, E. C., & Bartholow, B. D. (2017). The iterative nature of person construal: Evidence from event-related potentials. *Social Cognitive and Affective Neuroscience*, *12*(7), 1097–1107. https://doi.org/10.1093/scan/nsx048

Volpert-Esmond, H. I., Merkle, E. C., Levsen, M. P., Ito, T. A., & Bartholow, B. D. (2018). Using trial-level data and multilevel modeling to investigate within-task change in event-related potentials. *Psychophysiology*, *55*(5), Article e13044. https://doi.org/10.1111/psyp.13044

Von Gunten, C. D., Volpert-Esmond, H. I., & Bartholow, B. D. (2018). Temporal dynamics of reactive cognitive control as revealed by event-related brain potentials. *Psychophysiology*, *55*, Article e13007. https://doi.org/10.1111/psyp.13007

Vossen, H., Van Breukelen, G., Hermens, H., Van Os, J., & Lousberg, R. (2011). More potential in statistical analyses of event-related potentials: A mixed regression approach. *International Journal of Methods in Psychiatric Research*, *20*(3), e56–e68. https://doi.org/10.1002/mpr.348

Wagner, C. H. (1982). Simpson's paradox in real life. *The American Statistician*, *36*(1), 46–48. https://doi.org/10.1080/00031305.1982.10482778

Walker, P. M., Silvert, L., Hewstone, M., & Nobre, A. C. (2008). Social contact and other-race face processing in the human brain. *Social Cognitive and Affective Neuroscience*, *3*(1), 16–25. https://doi.org/10.1093/scan/nsm035

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience*, *9*(7), 971–978. https://doi.org/10.1038/nn1727

Wiese, H., Kaufmann, J. M., & Schweinberger, S. R. (2014). The neural signature of the own-race bias: Evidence from event-related potentials. *Cerebral Cortex*, *24*(3), 826–835. https://doi.org/10.1093/cercor/bhs369

Wijers, A. A., Mulder, G., Okita, T., Mulder, L. J. M., & Scheffers, M. K. (1989). Attention to color: An analysis of selection, controlled search, and motor activation, using event-related potentials. *Psychophysiology*, *26*(1), 89–109. https://doi.org/10.1111/j.1469-8986.1989.tb03137.x

Willadsen-Jensen, E. C., & Ito, T. A. (2006). Ambiguity and the timecourse of racial perception. *Social Cognition*, *24*(5), 580–606. https://doi.org/10.1521/soco.2006.24.5.580

Willadsen-Jensen, E. C., & Ito, T. A. (2008). A foot in both worlds: Asian Americans' perceptions of Asian, White, and racially ambiguous faces. *Group Processes & Intergroup Relations*, *11*(2), 182–200. https://doi.org/10.1177/1368430207088037

Willadsen-Jensen, E. C., & Ito, T. A. (2015). The effect of context on responses to racially ambiguous faces: Changes in perception and evaluation. *Social Cognitive and Affective Neuroscience*, *10*(7), 885–892. https://doi.org/10.1093/scan/nsu134

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*(8), 2031–2046. https://doi.org/10.3758/BF03196680

Woodworth, R. S. (1938). *Experimental psychology*. Holt.

Zebrowitz, L. A. (2006). Finally, faces find favor. *Social Cognition*, *24*(5), 657–701. https://doi.org/10.1521/soco.2006.24.5.657

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, *12*(4), 695–719. https://doi.org/10.1177/1094428108327450

Zhao, L., & Bentin, S. (2008). Own- and other-race categorization of faces by race, gender, and age. *Psychonomic Bulletin & Review*, *15*(6), 1093–1099. https://doi.org/10.3758/PBR.15.6.1093