2

PERCEIVING PERSONS

Social Cognitive Neuroscience Approaches

Curtis D. Von Gunten, Bruce D. Bartholow, and Hannah I. Volpert

Person perception is concerned with one person’s (i.e., a perceiver’s) attempts to understand what another person (i.e., a target) is really like inside (see Jones, 1990). Person perception is fundamental to virtually everything we do as a social species, in that our ability to engage with our world requires that we be able to understand efficiently and accurately whether others mean us harm, under what conditions we can expect their assistance, and whether our basic schemas for social interaction are appropriate. In this effort, perceivers are at the same time both optimistic and cynical—we often generally expect others to have positive dispositions (see Cacioppo, Gardner, & Berntson, 1997), while at the same time being dubious of others’ true nature, particularly if they initially confirm our positive expectations (Vonk & Van Knippenberg, 1994).

The act of person perception involves numerous, interrelated processes unfolding over time. It can be considered at the individual or group level. It can be influenced by a host of situational and dispositional factors. It is beyond the scope of this chapter to consider all such variables. Instead, we will restrict this review to research in which some aspects of person perception have been investigated using measures and methods derived from cognitive neuroscience, focusing in particular on research using event-related brain potentials (ERPs) and, to a lesser extent, functional magnetic resonance imaging (fMRI). Before proceeding, we first present a brief overview of the theory and methods of ERP and fMRI. Our focus on these two methods does not imply that other psychophysiological measures are not useful for the study of person perception; rather, these methods simply have been the most commonly used to date.
Neural Measures Used in Person Perception Research

Researchers interested in characterizing brain responses involved in a host of social psychological phenomena have a number of options available to them. Ideally, the decision concerning which measure(s) to use—and, critically, whether or not to use any measure of brain activity at all—is driven by the specific questions one wishes to address and the utility of such measures for addressing them (see Harmon-Jones & Beer, 2009). Indeed, it is important to keep in mind that nearly everything we know about person perception has been learned without the use of any neural measures whatsoever, and researchers would do well to consider whether their hypotheses can be addressed just as easily using behavioral and/or self-report measures alone, rather than complicating their design by adding a neural measure (see Amodio & Bartholow, 2011). Assuming a neural measure will add value to a given line of research, a number of additional considerations will help to determine which measure is most appropriate. For example, when a hypothesis concerns how quickly mental operations unfold, a measure that can adequately represent the mental operations in question in a temporally sensitive way should be chosen. The ERP is an excellent choice for addressing such questions.

The ERP is an electrical potential generated by the brain in response to a specific event, such as the presentation of a stimulus or the delivery of a response. Berger (1929) first demonstrated that it is possible to measure electrical activity generated from within the living human brain, a form of measurement known as the electroencephalogram (EEG). The continuous recording of EEG (e.g., during a psychological task) measures changes in patterns of brain voltage over time (for more information on the EEG, see Harmon-Jones & Peterson, 2009). When measured in the context of an experimental task involving specific stimulus and/or response events, it becomes possible to examine portions of the EEG that reflect neural responses uniquely associated with those events, at a millisecond (ms) level of temporal resolution. This event-related EEG response comprises the ERP.

Physiologically, ERPs represent the summation of post-synaptic electrical potentials generated by populations of synchronously active, primarily cortical neurons (see Coles & Rugg, 1995). Psychologically, ERPs represent information-processing operations instantiated by a stimulus or response event. The ERP waveform is composed of a series of positive and negative voltage deflections often referred to as components (see Figure 2.1). Although specific ERP components are often described as reflecting a particular information-processing operation, it is likely that any given ERP component represents numerous, simultaneously occurring processes (Fabiani, Gratton, & Federmeier, 2007). In general, the amplitude of a given ERP component represents the extent to which those operations are engaged by a stimulus or response event, and the latency of the component is thought to index the time needed to carry out those operations.
FIGURE 2.1 A Schematic Representation of an ERP Waveform Elicited by a Novel Visual Stimulus.

Note: The representation shows the waveform’s voltage (in microvolts; μV) by time (in ms) function. The vertical arrow on the timeline represents the moment of stimulus onset. The positive and negative deflections in the waveform represent typical ERP components, named here according to their polarity (”P” for positive and “N” for negative) and the approximate post-stimulus time of their peaks. Note, however, that this temporal naming convention is based on broad generalities and often does not conform to observed peak latencies. Note, also, that negative voltages are plotted “up” (i.e., above zero microvolts) on the Y axis, following electrophysiological convention.

The largest limitation to ERPs is their poor spatial resolution (though neural sources of ERPs can be estimated; see Huizenga & Molenaar, 1994). In contrast, fMRI is ideally suited to determining which neural structures are engaged by particular stimuli. As it is most commonly used, fMRI provides an indirect measure of the amount of neural activity occurring in various brain structures by measuring the increase in oxygenated blood flowing to those structures. Functional magnetic resonance imaging scanners contain powerful electromagnets that manipulate and detect differences in magnetic signals given off by oxygenated and deoxygenated hemoglobin carried in the blood, resulting in the blood-oxygen-level dependent (BOLD) signal. Areas with higher levels of oxygenated hemoglobin have a high BOLD signal, which is indicated by a brighter image, compared to areas with lower levels of oxygenated hemoglobin (Hyder, Shulman, & Rothman, 1998). Relevant to experimental psychology paradigms, the amount of oxygenated blood in an area of the brain and the rate at which it changes can be used as a corollary of the amount of neural activity occurring in that area, and therefore the extent to which that neural structure is involved in the information-processing operations elicited by the
experimental paradigm. Because of a delay in the hemodynamic response (about 4–6 seconds post-stimulus), fMRI is not very temporally precise and therefore is not useful for investigating temporal relationships among rapidly unfolding neurocognitive processes.

**ERP and fMRI Studies of Person Perception**

**Basic Processes**

Although there is no widely accepted method of categorizing and organizing psychological processes, here we have organized the mental operations of interest to person perception roughly in terms of their logical priority at first sight. We begin with face processing—a major determinant in recognizing an object as a person in the first place—and proceed through more “complex” processes, such as placing a person within socially relevant categories, inferring personality traits from behaviors, and the consequences of such processes for ongoing interactions. While levels of processing are at least partially separable at a conceptual level, it should be acknowledged that at an empirical level they may overlap. In fact, cognitive neuroscience provides an additional instrument for “lumping and splitting” (Boles, 2000) psychological processes, and as will become clear, has provided useful information regarding the separability of these processes, including the extent to which they share neurological substrates and occur in a causally linear sequence.

**Face Perception**

The most important source of information relevant for person perception is the human face. Faces occupy a unique status among social stimuli in that they rapidly convey large amounts of information relevant to social exchange, including a target’s specific identity, the social groups to which the target belongs (i.e., social categorization), cues to a target’s desirability for mating (i.e., attractiveness; general health), and what the target might be thinking or feeling (i.e., attention, motivation, and emotion). People need only milliseconds to extract this kind of critical social information from faces (see Ambady & Rosenthal, 1992) and to make decisions concerning social interaction, such as with whom to cooperate (Zebrowitz & Montepare, 2008), who is incompetent (Willis & Todorov, 2006), and who is likely to be aggressive (Bar, Neta, & Linz, 2006).

There is a vast literature on the cognitive neuroscience and neuropsychology of face perception and face processing (e.g., Atkinson & Adolphs, 2011; Haxby, Hoffman, & Gobbini, 2000), which we can mention here only briefly. Studies of face perception using ERP have identified a negative-going component, typically peaking around 170 ms following face onset and largest at lateral, occipito–temporal scalp locations, known as the N170 (see Figure 2.2). The N170 is
especially sensitive to faces relative to familiar non-face stimuli, such as hands, cars, and houses (Bentin & Deouell, 2000), and is believed to reflect early structural encoding of faces (Eimer, 2000). The N170 emanates from a network of posterior cortical regions near and including the fusiform gyrus (Deffke et al., 2007; Haxby et al., 2000). The N170 component is of special interest for person perception research because it presumably reflects the initial process of recognizing that an object is a conspecific (member of the same species) (Pascalis, de Haan, & Nelson, 2002)—the earliest stage of social perception. Despite relatively limited spatial specificity relative to fMRI, measuring the N170 has some notable advantages in that the temporal resolution of EEG can distinguish initial face encoding from later more elaborative processing. Moreover, recent evidence also suggests that the N170 is sensitive to higher-level social or motivational factors (Ofan, Rubin, & Amodio, 2014; Ratner & Amodio, 2013), suggesting that top-down processes play a role within 200 ms of seeing a face.

**Social Categorization**

The human mind seems spontaneously to place objects, including other people, into categories (Mervis & Rosch, 1981). It has long been assumed that people
are categorized for the same reason that anything else is—because categorization provides an efficient way to make available a host of expectations and assumptions regarding another person, thereby facilitating decision-making (see Fiske & Neuberg, 1990). Social categorization is thought to take place automatically, occurring without the perceiver’s intention or effort (see Brewer, 1988). In most behavioral research, categorization is inferred indirectly from various judgment biases that demonstrate the activation of stereotypes or affect linked to social categories (see Macrae & Bodenhausen, 2000), or from the speed with which various categories can be identified (see Zarate & Smith, 1990). Such research is limited by the fact that a single, static outcome (e.g., reaction time) is measured, which itself stems from multiple underlying mental operations that remain hidden.

Fortunately, ERPs provide an excellent tool for understanding social categorization as it unfolds over time, particularly in terms of the potential separability of sub-processes and their contributions to overt category judgments. For example, ERPs have been used to understand the so-called other race effect (ORE), in which individuals from the perceiver’s racial ingroup are recognized more quickly and accurately than individuals representing other racial categories (see Meissner & Brigham, 2001). One popular account of the ORE focuses on perceptual expertise. Since same-race (SR) faces typically are encountered more often than cross-race (CR) faces, perceivers theoretically have more expertise in processing faces from their own racial category. This frequency difference is thought to confer a configural or holistic processing advantage (i.e., processing emphasizing the relations among features of a stimulus; see Maurer, Le Grand, & Mondloch, 2002) for SR relative to CR faces, which must be processed using featural information (i.e., processing which emphasizes the individual features of a stimulus independently of their interrelationships; see Rhodes, Brake, Taylor, & Tan, 1989).

Given the N170’s association with configural or holistic processing, researchers have used it to examine the perceptual expertise account of the ORE. One approach has been to investigate the face inversion effect (FIE; see Yin, 1969) to SR and CR faces, which refers to the deficit in facial recognition for faces presented upside-down, compared to both other inverted objects and to non-inverted faces. Studies investigating upright and inverted SR and CR faces using N170 have shown a larger FIE for SR faces relative to CR faces (see Caharel et al., 2011). Moreover, Montalan et al. (2013) found that the N170 difference between SR and CR faces progressively diminished as a function of face rotation away from an upright orientation, such that, at full inversion (i.e., 180 degrees), when both SR and CR faces are expected to be processed in a feature-specific manner, the ORE was absent. These findings suggest that the ORE may be a product of SR faces being processed in a more configural manner than CR faces, generally supporting a perceptual expertise account of the ORE.
With regard to upright faces alone, N170 research using SR and CR faces has been inconsistent, with some researchers reporting an increase in N170 amplitude to SR faces (Ito & Urland, 2005), while others have reported an increase in the N170 to CR faces (Walker, Silvert, Hewstone, & Nobre, 2008) or no sensitivity of the N170 to race at all (Caldara, Rossion, Bovet, & Hauert, 2004). Some have attributed these variable results to differences in experimental task goals (e.g., individuating versus categorizing faces). To investigate this possibility, Senholzi and Ito (2013) examined the influence of race on the N170 by instructing participants to attend to the race of facial stimuli or to the unique identity of the faces. When participants attended to race, N170s were larger to SR than CR faces; but the opposite pattern emerged when participants attended to faces in an individuating way. In the individuating condition, the authors claimed that the enhanced N170 to CR faces reflects the additional demands placed on holistic information processing of those faces since CR faces are typically not encoded at the level of individuating features. In the categorizing condition, the enhanced N170 to same-race faces is explained by the default tendency to process SR faces in a holistic way. These results generally agree with an account of the N170 as a measure of the difficulty associated with processing a face in a holistic manner, with a larger N170 indicating greater difficulty.

However, the ORE also has been interpreted in the context of a social identity framework (Young, Hugenberg, Bernstein, & Sacco, 2012). Although perceivers generally have more expertise with members of their ingroups, decades of research on so-called “minimal groups” (Tajfel, Billig, Bundy, & Flament, 1971) and the tenets of Social Identity Theory (Tajfel & Turner, 1986) have shown that one need not have any history of association (and, hence, expertise) with members of a group in order to categorize them along the ingroup/outgroup dimension. Rather, the simple act of determining, for any number of even arbitrary reasons, that some individuals are “us” and others are “them” (i.e., assigning social identities) is sufficient to instigate various biases and even group-based discrimination. Thus, it could be that even the very earliest stages of face perception underlying the ORE (and its converse, the own-race bias) are subject to modulation by the motivational processes associated with identifying social groups (see Amodio, 2010).

Ratner and Amodio (2013) directly tested this possibility using a design in which both White and African-American perceivers were assigned arbitrarily to one of two groups—over-estimators or under-estimators—on the basis of their performance in a bogus numerical estimation task (see Tajfel et al., 1971). They then viewed faces of other “over-estimators” and “under-estimators” while ERPs were recorded. Critically, the race of the target faces was fully crossed with their group assignment, allowing Ratner and Amodio to assess the importance of social identity, independently of racial identity, in determining N170 amplitude. Consistent with prior behavioral (e.g., Young &
Hugenberg, 2010) and fMRI research (e.g., Van Bavel, Packer, & Cunningham, 2011) indicating that minimal group categorization enhances processing of ingroup faces, Ratner and Amodio found that ingroup targets were categorized more quickly at the behavioral level and elicited larger N170 amplitude than did outgroup targets, regardless of their racial group categories. Given that perceivers had essentially no experience in perceiving ingroup or outgroup targets prior to their arbitrary assignment to these groups, these data provide convincing evidence that amplitude differences in the N170 are driven in part by motivational factors associated with social identity, not just by perceptual expertise with ingroup targets (see also Zheng & Segalowitz, 2014).

Numerous other ERP components have been shown to be influenced by social group categorization and have been used to investigate the automaticity of social categorization. Mouchetant-Rostaing and colleagues (2000) investigated whether categorization by gender occurs spontaneously (i.e., even when perceivers’ overt goal is focused on a different task and with very limited visual information about targets). These researchers asked participants to view pictures containing either faces or body parts (hands and torsos) belonging to men, women, or both men and women (presented in different trial blocks). Participants performed either an explicit gender categorization task or a different task (e.g., scanning for other features, such as eyeglasses) while ERPs were recorded. Regardless of which task participants performed, a positive-going deflection in the ERP, beginning around 145 ms post-stimulus, differentiated the gender of the targets, an effect later replicated with faces varying according to age (Mouchetant-Rostaing & Giard, 2003). Moreover, this effect was restricted to trial blocks containing faces (i.e., it did not occur in blocks where only body parts were shown), supporting the importance of face processing for social categorization.

These initial studies were followed by numerous other ERP investigations of social categorization, including studies involving multiply categorizable targets. For example, Ito and Urland (2003) asked participants to view faces that varied by both gender (male and female) and race (Black and White) and asked them to categorize the faces in terms of either race or gender. This paradigm allowed for an examination of whether implicit categorization along one dimension occurs even while perceivers are explicitly categorizing faces along a different dimension. Ito and Urland reported neural differentiation along the racial dimension as early as the first negative-going deflection, peaking approximately 120 ms following onset of target faces, termed the N100, which was larger to Black than to White faces (in their predominantly White sample). The subsequent voltage deflection, a positivity dubbed the P200 with a peak latency around 180 ms, also was larger to Black than to White faces and additionally differentiated the gender of the faces, being larger to males than to females. Following the P200, the N200 (peaking around 260 ms in their study) was larger to females and Whites than to males and Blacks. Critically, all of these
effects occurred regardless of whether participants explicitly attended to race or
to gender, supporting the spontaneous nature of social categorization and
suggesting that early attention to these categories is determined more by the
properties of the target than by the overt goals of the perceiver.

However, some properties of the perceiver could importantly modulate early
stages of the categorization process, particularly with respect to the interaction
between perceivers’ and targets’ social categories. Ito and Urland (2003), for
example, used a participant sample that was mostly White (and included no
Black participants) and found that the P200 was larger to Black faces while the
N200 was larger to White faces. This pattern could emerge due to physical
features that distinguish Whites from Blacks, or to the spontaneous activation
of category-based information (i.e., stereotypes) that differs for Whites and
Blacks, in which case a similar pattern might be expected regardless of the race
of the perceivers. In contrast, such a pattern could result from motivational
processes associated with perceiving others as belonging to one’s ingroup versus
an outgroup. To test this latter hypothesis, Dickter and Bartholow (2007) had
both White and Black participants view faces of White and Black targets while
ERPs were recorded. Their findings replicated Ito and Urland’s findings (2003)
for White participants—a larger P200 to Black faces and a larger N200 to White
faces. However, among their Black participants, this pattern was reversed—a
larger P200 to White faces and a larger N200 to Black faces. This pattern
suggests that early attention to social category information depends not just on
the physical features or semantic knowledge associated with particular targets,
but is sensitive to distinguishing ingroup from outgroup members. Further
supporting the importance of this distinction, fMRI research has found that
both Blacks and Whites experience greater activity in the fusiform gyrus when
viewing pictures of racial ingroup members compared to racial outgroup
members (Golby, Gabrieli, Chiao, & Eberhardt, 2001).

Additional evidence of the role of motivational processes in modulating
early perceptual attentional processing of outgroup faces was provided by
Amodio (2010), who examined whether individual differences in approach
motivational states, as indexed by frontal cortical asymmetry (see Harmon-
Jones, 2003), altered perceptual attention to racial ingroup and outgroup faces.
Amodio reported that greater approach motivation (i.e., stronger relative left-
frontal cortical activity) among the White participants in his experiment
predicted enhanced amplitude of the P200 elicited by Black faces, which in
turn predicted less biased performance on a measure of implicit racial bias.

**Summary**

The pattern of results reported by Amodio (2010) provides support for the idea
that motivation serves a functional role in person perception (Bruner, 1957)—in
this case, to tune perception according to currently active goals in order to
facilitate goal-direction action. More generally, this review highlights that recent social neuroscience approaches to person perception are helping to redefine traditional notions of the role of social category information in determining prejudice-related responses. Specifically, traditional dual-process models of person perception (e.g., Brewer, 1988; Devine, 1989) held that initial perceptual and early attentional processes are reflexive, leading inevitably to automatic biases unless interrupted by more deliberative, reflective (i.e., controlled) processes that are subject to a perceiver’s goals and motivations. Early social neuroscience research appeared to support such a dual-process framework by suggesting that the neural substrates supporting reflexive and reflective processes were separable. For example, Phelps et al. (2000) found that amygdala activation to race information was strongly correlated with implicit, but not explicit, attitudes, while Ito and Urland (2003) found that the P300 elicited by race was correlated with explicit, but not implicit, measures of racial bias. Moreover, activation in the anterior cingulate cortex (Amadio et al., 2004) and lateral prefrontal cortex (Cunningham et al., 2004), areas associated with self-regulatory cognitive control (see Braver, 2012), were found to increase when people had the motivation and opportunity to control racially biased responses.

However, the findings reported in more recent studies, as reviewed in this section, have challenged a basic premise of the dual-process framework by showing that even the earliest neural responses to social category information are subject to modulation by motivational factors. Such findings suggest that person perception, similar to other, more basic aspects of information processing (see Coles, Gratton, Bashore, Eriksen, & Donchin, 1985), probably unfolds not in a discrete, stage-like manner, but rather as a continuous process in which both bottom-up (i.e., stimulus-driven) and top-down (i.e., goal-driven) processes jointly influence the creation of “person concepts” (see Freeman & Ambady, 2011; Srull & Wyer, 1989).

More Complex Processes

Trait Inferences and Expectancy Violation

We now turn from the most basic “perception” side of the person perception equation to the more complex “person” side, examining research aimed at understanding how perceivers appear to utilize information to form ideas, expectations, and judgments about the people they encounter. In particular, the process of inferring a target person’s traits from her behaviors and integrating those traits into a cohesive impression of a person is an essential aspect of person perception (Asch, 1946). Research has shown that people extract surprising amounts of information about a person’s nature and personality from very little contact (Ambady & Rosenthal, 1992). Essentially, the primary purpose of person perception is to create expectancies about other people’s behavior,
which then guide our decisions and behaviors with respect to those others in an effective way (Jones, 1990; Roese & Sherman, 2007).

Through years of learning what to expect from people given their presumed traits, our cognitive representations of others are tightly bound to our expectancies (Roese & Sherman, 2007). The importance of such expectancies for ongoing social interactions has especially been demonstrated through research investigating how people react when others’ behaviors violate our expectations. In essence, expectancy violations are motivationally significant events because they alert us to the fact that our current understanding of a person or a social situation is at least partly inaccurate, which has implications for our own future behavior. In the laboratory, behavioral (e.g., recall) and neural responses to expectancy-violating information can help researchers to draw inferences about the properties of initial expectancies, such as their strength, representation in different neurocognitive systems, and likelihood of modification in the face of new information.

The neurocognitive processes instantiated by expectancy violations largely have been investigated using ERPs, mainly because researchers often are interested in characterizing processes thought to unfold very rapidly (see Olson, Roese, & Zanna, 1996). In particular, researchers have investigated the effects of expected and unexpected person information on the amplitude of the P300 component. Several theories of the P300 have been proposed, mainly centering on its sensitivity to novelty (Friedman, Cycowicz, & Gaeta, 2001) and the subjective probability of eliciting events (Duncan-Johnson & Donchin, 1977). As applied to person perception, the most relevant ideas about the P300 link its amplitude to context updating in working memory (Donchin & Coles, 1988) and more generally to the motivational significance or relevance of an eliciting stimulus (see Nieuwenhuis, Aston-Jones, & Cohen, 2005).

The P300 was so named because in early studies (mainly investigating processing of novel auditory stimuli), the component’s peak latency occurred reliably around 300 ms post-stimulus (see Sutton, Braren, Zubin, & John, 1965). However, it is not uncommon for the P300 to peak substantially later than 300 ms in tasks involving complex social or emotional stimuli (see Bartholow & Amodio, 2009; Bartholow & Dickter, 2007), a fact that has led some to posit a difference between the “classic” P300 elicited in purely cognitive tasks and a “late positive potential,” or LPP, elicited by stimuli that carry more social or emotional significance (e.g., Schupp et al., 2000).

The first study specifically to investigate person perception processes using a neurocognitive measure was reported by Cacioppo, Crites, Gardner, and Berntson (1994), who found that personality trait words that were inconsistent with a context established by previous trait words elicited enhanced P300 amplitude. In another early experiment, Osterhout, Bersick, and McLaughlin (1997) showed that a P300-like late positivity in the ERP waveform was sensitive to violations of gender stereotypical noun–pronoun agreement in
sentence comprehension (e.g., “The surgeon prepared herself for the operation.”).
This effect was independent of participants’ self-reported judgments of the
acceptability of the sentences, suggesting that the P300 might provide a
relatively covert indication of implicit person judgments (also see Crites,
Cacioppo, Gardner, & Berntson, 1995).
Based in part on this earlier study, Bartholow, Fabiani, Gratton, and
Bettencourt (2001) reasoned that processing of interpersonal expectancy violations
(as an index of a prior trait inference) also should manifest in P300 amplitude.
Numerous studies in social and developmental psychology indicate that
expectancy-violating information about people often is recalled better than
expectancy-consistent information (see Stangor & McMillan, 1992). Theoretical
models (e.g., Srull & Wyer, 1989) posit that this recall advantage reflects updating
of working memory that occurs during inconsistency resolution, the process by
which people attempt to reconcile the discrepancy between new information
and existing person concepts, an idea highly relevant to the context updating
theory of the P300.
Bartholow and colleagues (2001; see also Bartholow, Pearson, Gratton, &
Fabiani, 2003) tested this idea in a set of experiments in which participants read
paragraph descriptions of several fictitious individuals in order to form
impressions of them, and then read sentences depicting behaviors that were
either consistent or inconsistent with those impressions. In line with the
context updating hypothesis, P300 amplitude elicited by expectancy-violating
behaviors was larger than that elicited by expectancy-consistent behaviors (also
see Van Duynslaeger, Sterken, Van Overwalle, & Verstraeten, 2008). Expectancy-
violating behaviors also were better recalled than expectancy-consistent behaviors,
further supporting the notion that P300 amplitude reflects the extent to which
stimuli are processed during memory encoding (see Fabiani & Donchin, 1995).
Findings from these studies suggest that violations of expectancies derived
from a given individual’s behavior prompt cognitive operations reflected in
the P300, which in theory represent the updating of templates in working
memory (e.g., Donchin & Coles, 1988) and motivated attention to information
relevant to ongoing interpersonal interactions (Nieuwenhuis et al., 2005).
Interestingly, individual target-based expectancy violations have not been
shown to influence the N400 component of the ERP. The N400 is a negative-
going voltage deflection apparent roughly 400–600 ms following onset of
a stimulus incongruent with a semantic context. In a now classic example,
Kutas and Hillyard (1980) showed that the final word of the sentence, “The
pizza was too hot to cry” elicits a much larger N400 than does the final word
of the semantically more sensible sentence, “The pizza was too hot to eat.”
In at least one previous study (Bartholow et al., 2001), the paradigm was
designed explicitly to compare effects of trait-based expectancy violations and
semantic violations, and N400 effects were observed only for the latter. This
suggests that trait-based expectancies are not held in memory as semantic
knowledge, but perhaps are closer to episodic memory traces (see Jones & McGillis, 1976).

In contrast, however, expectancies derived from targets’ memberships in social categories do seem to be represented in memory as semantic knowledge. Such category-based expectancies are essentially equivalent to stereotypes, which researchers have long assumed to be similar to other forms of category-based knowledge held in semantic memory (see Hamilton, 1981; but see Contreras, Banaji, & Mitchell, 2011). Several studies have shown N400 responses to stereotype violations. For instance, Van Berkum and colleagues (2008) reported enhanced N400 amplitude when perceivers heard a male voice stating preferences that violate male gender norms (e.g., “I wish I looked like Britney Spears.”). Similarly, enhanced N400 amplitude has been reported when gender-based expectancies are primed using simple category labels (e.g., “Women” or “Men”) followed by trait words that violate gender stereotypes (e.g., “nurturing” or “aggressive,” respectively) (see White, Crites, Taylor, & Corral, 2009). Violation of expectancies based on racial and ethnic categories also has been shown to elicit enhanced N400 amplitude (Hehman, Volpert, & Simons, 2014; Wang et al., 2011). Note, however, that at least two studies examining category-based expectancy violation have failed to show evidence of enhanced N400 amplitude to such violations (Bartholow, Dickter, & Sestir, 2006; Osterhout et al., 1997). Perhaps most surprisingly, Osterhout et al.’s (1997) study, described previously in this section, used verbal stimuli that largely mimicked the typical semantic violation paradigm (e.g., “The beautician put himself through school.”), and yet gender category violations elicited a P300-like response and not an N400.

Valence Processing

In addition to whether or not others’ behaviors are consistent with our expectations, the valence of those behaviors—their positivity or negativity—appears very important for shaping both information processing and impression formation. It has long been observed that negative behaviors have a stronger influence on person perception than positive behaviors (e.g., Asch, 1946)—the so-called positive–negative asymmetry (Peeters & Czapinski, 1990). Of importance for the current review, some evidence indicates that valence and expectancy interact to determine the processing of target behaviors. In particular, when others’ behavior is both negative and unexpected, it appears to receive particularly enhanced processing. Ybarra, Schaberg, and Keiper (1999) demonstrated this effect using memory measures, reporting that negative expectancy-violating behavior tends to be remembered better than positive expectancy-violating information and negative expectancy-consistent information. To explain this pattern, Ybarra (2002) proposed a naïve causal theories model, which posits that positive behaviors lead to greater uncertainty in trait inferences because such behaviors are most often attributed to situational pressures that encourage positive behavior in
social settings. Negative behaviors, on the other hand, are not socially encouraged and therefore are more readily attributed to the actor’s disposition. According to this model, because initial positive impressions are less certain, they are more modifiable on the basis of new and inconsistent information, which, when encountered, is processed more extensively as perceivers attempt to integrate it into their existing impression of the target. In contrast, impressions formed on the basis of negative behaviors tend to be more certain and should be less easily modified by subsequent information (see Vonk & Van Knippenberg, 1994). In other words, effortful inconsistency resolution is more likely with unexpected negative behaviors than with unexpected positive behaviors.

This model has been supported by ERP studies showing that negative information about initially positive targets elicits a larger P300 than positive information about initially negative targets (Bartholow et al., 2003; Cacioppo et al., 1994). Moreover, recent work has demonstrated that this differentiation occurs even earlier in the information-processing stream, such that negative

---

**FIGURE 2.3** ERPs, Measured at a Right-lateralized Frontal Electrode (F8), Elicited by Words Depicting Positive and Negative Behaviors that either Confirmed (Expectancy-consistent; EC) or Violated (Expectancy-violating; EV) Previously Established Trait-based Expectancies.

*Note:* Negative behaviors that violated previously established (positive) expectancies elicited larger amplitude of the P200, one of several early-latency deflections often associated with visual attention.
expectancy-violating behaviors elicit larger amplitude of an early attention-
related component, the P200, than do positive expectancy-violating or negative
expectancy-consistent behaviors (Jerónimo, Volpert, & Bartholow, 2015; see
Figure 2.3), presumably reflecting attention directed spontaneously to information
that should receive more elaborative processing (i.e., unexpected, negative
information).

Causal Attributions

The preceding discussion underscores the importance of causal attribution in
person perception. Causal attribution refers to the act of explaining the cause
of a person's behavior in terms of either stable internal traits or as the result of
fleeting, situational events. Perceivers very frequently use targets' behaviors to
infer their traits. The positive–negative asymmetry, mentioned in the previous
section, is often explained in terms of the diagnosticity of negative versus
positive behaviors for trait inferences (Reeder & Brewer, 1979). In essence,
negative actions are more diagnostic because, in theory, they can only be
performed by immoral actors, whereas positive actions can be performed by
both moral and immoral actors (Skowronski & Carlston, 1989). In other words,
negative behaviors often are attributed to internal characteristics of the actor,
whereas positive behaviors are more easily attributed to situational factors.

Researchers have long observed that perceivers generally downplay the role
of situational factors when attributing the causes of others' behavior (Gilbert
& Malone, 1995; Ross, 1977). For instance, an explanation of someone's
expertise at an activity (e.g., chess or soccer) may rely too heavily on an
assumption of natural talent while downplaying the role of practice, or even
essentially arbitrary factors such as their birth month (see Dudink, 1994). This
tendency—often referred to as the correspondence bias—is associated with the
psychological predisposition to form impressions of others. The natural
proclivity to attribute mental states to others, and to use that information to
predict future mental states and behavior, is itself a burgeoning area of research
(see Moran and Mitchell, Chapter 4 in this volume, for a review).

Research examining the neural foundations of trait inference points to the
involvement of the temporal–parietal junction (TPJ) and medial prefrontal
cortex (mPFC) in making inferences about a person's traits from their behavior
(e.g., Saxe & Powell, 2006). These same areas also have been implicated in
research investigating how perceivers detect others' goals from their behavior
(see Frith, U. & Frith, C., 2001) and how others' beliefs are represented
(Apperly, Samson, Chiavarino, & Humphreys, 2004).

Given the role these regions play in trait inference and mentalizing, Harris,
Todorov and Fiske (2005) adapted a widely used behavioral paradigm for
examining dispositional versus situational attribution to an fMRI context.
Behavioral studies generally show that three different sources of information
each influence how likely a person is to attribute dispositional over situational causes to a target's behavior (McArthur, 1972; Pruitt & Insko, 1980). More specifically, when told that Sue is afraid of dogs, participants are more likely to attribute Sue's emotional response to Sue's internal disposition than to the dog or other circumstances if (a) hardly anyone else is afraid of the dog (consensus), (b) Sue is also afraid of other dogs (distinctiveness), and (c) Sue regularly is afraid of that dog (consistency). While in the scanner, participants were shown a lead sentence such as “Sue is afraid of dogs,” after which a second screen presented the lead sentence again along with three other sentences used to manipulate consensus, distinctiveness, and consistency information. Finally a response screen appeared asking participants to choose one of four causes for Sue’s behavior: (a) Sue (person), (b) the dog (stimulus), (c) circumstance, (d) combination. Consistent with the role that the TPJ and mPFC play in psychological processes that involve representing a social target’s mental states, Harris et al. (2005) found that these two regions (along with the posterior superior temporal sulcus, or pSTS) were activated for dispositional inferences, but not for situational inferences. Further, Mitchell, Heatherton, and Macrae (2002) compared trait inferences about persons to property inferences about inanimate objects and found activation in similar mentalizing regions for trait inferences, but not for object property inferences.

Other fMRI work also has contributed to explaining the ease and spontaneity of trait inference. When not engaged in an active task, the brain is thought to reach a baseline or resting state (Gusnard & Raichle, 2001). However, it has been observed that brain regions differ in how much glucose and oxygen they metabolize while in this state. Importantly, several of the areas that show heightened baseline activity include the areas just discussed that are regularly implicated in mentalizing, such as ventral and dorsal mPFC, pSTS, TPJ, and precuneus (Gusnard, Akbudak, Shulman, & Raichle, 2001). The observation that the regions implicated in attributing mental states to social targets are also some of the same regions that exhibit the highest resting metabolic rate may explain the ease and spontaneity of trait inference, and people’s general propensity to rely too heavily on dispositional explanations of human behavior (see Lieberman, 2007). These findings also have been used to explain the ease with which humans anthropomorphize and personify inorganic, inanimate objects (Mitchell, 2006). For instance, research has found that intentions and motives can be attributed to animate geometrical objects (Heider & Simmel, 1944); and dispositional causes, rather than situational causes, can be attributed to objects such as pens and brooms (Harris & Fiske, 2008).

Summary

We have touched briefly on several of the most enduring topics in person perception (trait inferences and expectancy violations; valence processing and
causal attribution), and discussed how research on those topics has been advanced in recent years through the use of psychophysiological methods. Two general conclusions are important to highlight from this section. First, neural responses differentiating expected from unexpected behavior, and pleasant from unpleasant behavior (as well as the interaction of these two factors) occur very rapidly, as demonstrated by numerous ERP studies (e.g., Bartholow et al., 2003; Cacioppo et al., 1994). This finding is important primarily because it addresses a theoretical question concerning the speed with which such information can be understood and incorporated into existing person templates (Srull & Wyer, 1989), as well as whether valence information matters for determining such processing at very early stages (see Hamilton, Driscoll, & Worth, 1989). Second, evidence from fMRI and ERP source localization research is converging on the idea that a number of social cognitive processes share a neural network involving regions often implicated in “mentalizing” (Lieberman, 2007), including the mPFC, TPJ, precuneus, and pSTS (see Kestemont, Vandekerckhove, Ma, Van Hoesick, & Van Overwalle, 2013). This remarkable convergence across theoretically distinct aspects of person perception, along with the implication of these same areas in the brain’s so-called “default mode network” (see Raichle et al., 2001), points to the centrality of socially motivated cognition for human life and brain function (see Lieberman, 2007).

Conclusion

Since their introduction to the field in the mid-1990s, application of the methods and theory of cognitive neuroscience has dramatically advanced understanding of a number of aspects of person perception. (Note, too, that psychophysiological methods were used in a handful of studies on related topics, such as intergroup attitudes, as early as the 1950s; see Ito & Bartholow, 2009.). Perhaps one of the main methodological contributions permitted by the use of such measures is researchers’ ability to make inferences about the effects of experimental manipulations on psychological processes without requiring a behavioral response. This approach provides a way to circumvent participants’ attempts to control their responses in socially desirable ways, and permits a more direct assessment of the neural responses that underlie abstract concepts like “attitudes” and “inferences,” and their behavioral manifestations. Event-related brain potentials (ERPs), for example, allow researchers to separate the timing and contribution of underlying cognitive processes that contribute to overt responses, as well as processes associated with cognitive processing versus response implementation.

As with any new methodology, however, caution is needed in determining when and how to use these measures. Several authors have written informative pieces urging researchers to maintain rigorous standards for drawing appropriate inferences from neural and other physiological measures (see Cacioppo et al.,
Also, given the burdens they impose in terms of person hours (for both researchers and participants), modifications to standard paradigms, and financial costs, others (e.g., Amodio & Bartholow, 2011) have argued for limiting the use of such measures to situations and research questions when behavioral measures alone cannot provide adequate answers. In other words, theory-derived research questions should be used to determine which measures are appropriate, and not the other way around.

References


