1 Interpreting Event-Related Brain Potentials

Leun J. Otten and Michael D. Rugg

Our ability to feel, think, and act can in some way be attributed to the workings of the brain. For over a century, scientists have used measures of brain activity to gain insights into perceptual, cognitive, and motor functions. As a result, researchers have developed a variety of methods to measure brain activity noninvasively (e.g., Rugg, 1999). These methods roughly fall into two classes: “electromagnetic” approaches that directly measure brain activity by recording the electromagnetic fields generated by certain neuronal populations, and “hemodynamic” approaches that indirectly measure brain activity by recording changes in vascular variables that are linked to changes in neural activity. Importantly, these methods differ in a number of aspects, including the preconditions for detecting a signal, the homogeneity with which neural activity is sampled from different parts of the brain, and the relative strengths in determining when versus where neural activity takes place. They therefore provide complementary views on neural activity.

This chapter focuses on electromagnetic measures of neural activity. Within this class of methods, there are several ways to examine electrical and magnetic activity, in both the temporal and spatial domains (e.g., Näätänen, Ilmoniemi, & Alho, 1994; Tallon-Baudry & Bertrand, 1999). Here we restrict discussion to event-related brain potentials (ERPs), which are small changes in the electrical activity of the brain that are recorded from the scalp and that are brought about by some external or internal event (see Coles & Rugg, 1995; Kutas & Dale, 1997). This electrical activity changes rapidly over time and has a spatially extended field. It is therefore usually recorded with a temporal resolution in the order of a few milliseconds from multiple scalp locations. The goal of this chapter is to explain how one can make functional interpretations from ERP data. After a brief introduction to the issues that ERP analysis aims to address, we outline the type of inferences that one can and cannot make from ERP data. The final two sections then examine the assumptions that underlie functional inferences, and how functional interpretations of ERP data may develop in future. The material considered here is similar to that covered by Kutas and Dale (1997) and Rugg and Coles (1995).
What Issues Can ERP Analysis Address?

A first step toward making functional interpretations from ERP data is to consider what purpose ERPs serve. One can study ERPs in their own right, that is, to gain a better understanding of aspects of ERPs themselves. For example, there has been substantial work to characterize individual features of ERP waveforms, and to identify the intracerebral origins of ERPs. More often, however, researchers use ERPs as a tool to resolve questions in disciplines such as psychology, psychiatry, and neuroscience. For example, ERPs have helped to delineate psychiatric and neurological conditions such as schizophrenia and ADHD (e.g., Ford et al., 1999; van der Stelt et al., 2001), why people take longer to respond in situations of conflicting information (e.g., Duncan-Johnson & Kopell, 1981), how attention normally works (e.g., Mangun & Hillyard, 1995), and why memory declines as we grow older (e.g., Rugg & Morcom, in press). Attempts have even been made to use ERPs as a lie-detection tool (Farwell & Donchin, 1991)!

In this chapter, we confine our discussion of functional interpretations from ERPs to their use in the field of cognitive neuroscience, although the logic and assumptions laid out here also apply to most other applications. Cognitive neuroscience “aims to understand how cognitive functions, and their manifestations in behavior and subjective experience, arise from the activity of the brain” (Rugg, 1997, 1). We focus on what ERPs can reveal about cognitive functions in healthy individuals, using within-group comparisons. Comparisons between groups of individuals, especially when special populations such as clinical or younger/older people are involved, require additional considerations (see Picton et al., 2000; or Rugg & Morcom, in press, for introductions to this topic).

Explanations in cognitive neuroscience can be articulated at many different levels, ranging from functional to cellular and even subcellular accounts (e.g., Marr, 1982). One can use ERPs to address questions at several of these levels. For example, at a functional level, some use ERPs to address whether the brain honors the distinction between syntax and semantics (e.g., Friederici, 1995). At a lower level, researchers use ERPs to investigate the speed of interhemispheric transmission (e.g., Lines, Rugg, & Milner, 1984), or the effects of pharmacological manipulations (e.g., Hsu et al., 2003). Often, interest spans across levels, and explanations at one level may constrain explanations at another level. In the next section, we discuss how one can use ERP data to make functional inferences.

Making Inferences from ERPs

We can classify inferences from ERP data in several ways. It is possible to order inferences on the basis of their complexity and underlying assumptions (Rugg & Coles,
1995), or on the emphasis placed on the temporal versus the spatial information that ERPs provide. Here, we draw a distinction between inferences that one can make with and without adopting a functional interpretation of some feature of an ERP waveform. ERPs have been in use since the 1960s, and many studies have attempted to associate particular features of ERP waveforms with specific cognitive processes. On the basis of the findings of such studies, it is sometimes possible to use specific ERP features (or “components”—see below) as markers for the engagement of the cognitive process with which they are correlated. One can also draw meaningful interpretations of ERP data without making assumptions about the functional significance of any particular waveform feature. In the following sections, we therefore distinguish between inferences made with and without such theoretical commitments. We discuss the latter class of interpretation first.

**Inferences Not Based on Prior Knowledge**

ERPs can be employed to study cognitive processes even when there is little or no prior useful information to bring to bear on the functional significance of any feature of the elicited ERP waveforms. In practice, this is a common situation. There are generally three kinds of inferences made in these circumstances: about the timing, degree of engagement, and functional equivalence of the underlying cognitive processes. These inferences rely on three aspects of ERP differences observed between conditions: their time course, amplitude, and distribution across the scalp, respectively. We will illustrate these inferences with a concrete example.

Consider an experiment in which ERPs are elicited at three electrode sites in two conditions (1 and 2), and in two situations (A and B; see figure 1.1). The simplest type of inference from these data is based on the observation that the ERP waveforms elicited in the two conditions differ. (For this and all subsequent types of inference, this observation can be substantiated by an appropriate quantification of the waveforms; see chapter 3 of this volume). On the assumption that specific cognitive processes are manifested in specific and invariant patterns of neural activity (see below), a reliable ERP difference between conditions implies that the cognitive processes associated with the two conditions differ in some respect. Understanding how the cognitive processes differ depends on a conceptual analysis of the differences between conditions.

Even this simple inference can lead to useful insights. For example, a longstanding question in cognitive psychology is the level to which unattended information is processed. One can address this question by recording ERPs for unattended information, and establishing whether the content of the unattended information influences the ERP waveforms. Using this logic, researchers have found that ERP waveforms for unattended information differ when an unattended, visually presented word is presented twice in succession (Otten, Rugg, & Doyle, 1993). This suggests that unattended visual information can be processed to the level of its identity.
Expanding on the first type of inference, the second type of inference takes advantage of the high temporal resolution of ERP waveforms, which makes them especially valuable for drawing inferences about the timing of cognitive processes. In situation A of figure 1.1, the ERP waveforms in the two conditions start to differ at about 250 ms after the onset of the event of interest. This implies that the cognitive processes that differentiate the two conditions began to differ by 250 ms. Using this logic, researchers have demonstrated that the ERP waveforms elicited by attended and unattended stimuli can differ as early as 50 ms after stimulus onset (Woldorff & Hillyard, 1991). Accordingly, attentional processes must have been engaged within 50 ms, providing important information about the functional characteristics of selective attention.

The final two classes of inference discussed in this section are based on interpretation of the scalp distribution and amplitude of an ERP effect, respectively. Information about the scalp distribution of an ERP effect forms the basis of efforts to estimate the nature of the intracerebral sources that underlie the effect (e.g., Scherg, 1990). More importantly in the present context, however, this information contributes to the determination of whether functionally nonequivalent processes are engaged across conditions. Crucially, one can make such inferences even in the absence of knowledge about the intracerebral sources of the ERP effects in question.

In situation A of figure 1.1, the difference between the two conditions is largest at the parietal electrode site. By contrast, in situation B, the difference between conditions

**Figure 1.1**
Hypothetical ERP waveforms elicited at three electrode sites in two experimental conditions in two experimental situations (A and B). The differences between the waveforms allow a number of functional interpretations. See text for details.
is largest at the frontal electrode site. As we discuss later, there are several reasons why scalp distributions may change. Regardless of the cause, however, different scalp distributions imply that different patterns of neural activity are associated with the two situations. So far as one is willing to accept the assumption that experimental conditions that are neurophysiologically dissociable are most likely functionally dissociable as well (see below), one can use ERPs to assess whether the cognitive processes engaged in different experimental conditions are functionally distinct.

We can apply the same logic to differences in scalp distribution that emerge over time. ERP effects can be compared not only across experimental conditions as exemplified above, but also across time points within a condition, or across time points across conditions. In any case, a difference in scalp distribution implies a difference in underlying neural pattern. In turn, different neural patterns imply that distinct functional processes were engaged across conditions, times, or both.

For example, when people are asked to decide whether or not they remember having experienced an item before, new and old items elicit different ERP waveforms. This difference is largest over left parietal scalp sites in an early time region of the waveforms, before becoming largest over right frontal scalp sites later on (see Rugg & Wilding, 2000, for review). These scalp distribution differences suggest that different patterns of neural activity are engaged over time. Accordingly, memory retrieval may rely on multiple, qualitatively different functional processes, operating at different points in time. Without evidence that the two effects are dissociable, however, the possibility that they act in concert to support a common process cannot be ruled out. As it happens, the left parietal and right frontal ERP effects are sensitive to distinct experimental manipulations (Rugg & Wilding, 2000).

If scalp distributions do not differ across conditions or time, does this have any functional implications? If experimental manipulations do not result in scalp distribution differences, but the associated ERP effects nonetheless differ in amplitude, this is usually taken to suggest a quantitative, as opposed to a qualitative, difference in the cognitive processing engaged in the two conditions. That is, the experimental manipulations are thought to have engaged the same cognitive process(es), but to differing degrees. Later on, we discuss caveats surrounding interpretations from such amplitude differences and null results.

**Inferences Based on Prior Knowledge: ERP “Components”**

As discussed in the previous section, we can make useful inferences about cognitive processes from ERP data without knowing what any particular waveform feature represents. However, we can gain additional information with knowledge about the functional significance of some aspect of an ERP waveform. ERPs can be thought of as time-varying scalp fields that result from the summation of electromagnetic activity generated by neuronal populations in different parts of the brain. Clearly, it would be informative to understand these fields both in terms of the neuronal populations...
responsible for them and the different cognitive processes with which they are associated. In essence, this is what the decomposition of ERP waveforms in terms of their underlying “components” attempts to achieve.

There is no universally accepted definition of what constitutes an ERP component. Because neural and cognitive processes overlap in both space and time, features of the waveform such as peaks or troughs can result from the summation of several contributing sources, and thus may not reflect functionally homogeneous neural or cognitive processes. Component definitions range between two extremes, sometimes referred to as the “physiological” and “functional” approaches to component identification. According to the physiological approach (e.g., Näätänen & Picton, 1987), an ERP component should be defined in terms of its anatomical source within the brain. To measure a component, it is therefore necessary to isolate the intracerebral sources underlying an ERP waveform. By contrast, according to the functional approach (e.g., Donchin, 1981), an ERP component should be defined predominantly in terms of the functional process with which it is associated. On this account, it is irrelevant whether one or several anatomical sources contribute to the component, as long as they constitute a functionally homogeneous system.

In practice, ERP components are usually defined with respect to both their functional significance and their underlying neural source(s). Along these lines, Donchin, Ritter, and McCallum (1978) give an operational definition of an ERP component. According to this view, a component is a part of the waveform with a circumscribed scalp distribution (alluding to the underlying neural configuration) and a circumscribed relationship to experimental variables (alluding to the cognitive function served by the activity of this configuration). Several procedures, based on the analysis of scalp distribution and sensitivity to experimental manipulations, have been proposed as methods to dissociate and measure overlapping components (see Picton et al., 2000).

What can we gain from the concept of an ERP component? Despite the difficulties surrounding their definition and measurement, components serve at least three purposes. First, they provide a language that allows communication across experiments, paradigms, and scientific fields. Second, they can provide a basis for integrating ERP data with other measures of brain activity. Third, components can serve as physiological markers for specific cognitive processes. In the case of some components, sufficient information has accumulated to indicate, in broad terms at least, their functional significance. Below, we illustrate how one can make functional interpretations from ERP data using the notion of a component (see also chapter 2 of this volume).

Again, consider the waveforms illustrated in figure 1.1. Assume that in situation A, the positive deflection in the waveforms (labeled X and X’) is a known ERP component, associated with some specific cognitive process. (This assumption is for the purposes of exposition only. In reality, it is highly unlikely that such a large, temporally extended ERP deflection would reflect the activity of a single generator system, or a
single cognitive process.) On this assumption, the first inference one can make from these data makes use of the time course of the component across conditions. The time course can be quantified with one of several temporal measures of the component, for example its onset, peak latency, rise time, or duration (see chapter 3 of this volume). In figure 1.1, the component onsets later in condition 2 than 1. This implies that the cognitive process presumed to be associated with the component is engaged at a later time in condition 2 than 1.

Next, figure 1.1 shows that the amplitude of the component in situation A differs between conditions. As with the time course of a component, we can define its amplitude in several ways. The observed amplitude difference in figure 1.1 implies that the cognitive process is engaged to a different degree across conditions. This inference relies crucially on previous work associating variance in the amplitude of the component with variance in the degree to which the associated cognitive process is engaged. To illustrate this type of inference, based on its scalp distribution and approximate time of occurrence, the positive peak seen in situation A of figure 1.1 may reflect the P300 or P3b component (Donchin, 1981; Donchin & Coles, 1988). Donchin and Coles (1988) proposed that P300 amplitude variations reflect variations in the degree to which an internal representation of the experimental context is updated. On this account, the differences between conditions shown in situation A of figure 1.1 support the inference that updating processes are greater in condition 1 than 2. Such inferences based on amplitude measures only apply when comparing the same component across conditions.

The reader may have noticed that all the inferences discussed in this and the previous sections were framed in terms of comparisons between experimental conditions. That is, they are based on an analysis of differential ERP effects. Functional interpretations of any measure of neural activity rely crucially on a carefully designed experiment. The processes of interest must be isolated with judiciously selected experimental conditions. Virtually without exception, this requires the researcher to manipulate the process across two or more experimental conditions. Accordingly, functional interpretations are usually made from differences in neural activity, computed between the conditions that are presumed to isolate the process(es) of interest.

**What Cannot Be Inferred from ERPs?**

ERP data can provide valuable information about cognitive functions in many situations. When using ERP data to make functional interpretations, it is important to keep these strengths in mind. Equally important, however, is to recognize the limitations of ERP data. For example, ERPs can provide no information about neural activity giving rise to “closed” electromagnetic fields (see below). In addition, many of the inferences discussed in the previous sections rely on assumptions that may be violated in any
given case. In this section, we outline some of these issues. Note that most of these issues are equally relevant to noninvasive measures of brain activity other than ERPs.

**Null Results**
The first issue concerns null results. Several of the inferences discussed above follow from the finding that a comparison of interest failed to result in a statistically reliable difference in amplitude, scalp distribution, or latency. That is, the inference is based on the lack of an effect. For example, we described how the lack of a reliable scalp distribution difference across experimental manipulations may suggest that the same cognitive process is associated with each manipulation. However, for at least three reasons, interpretations based on the absence of an effect should be treated with caution. First, the experiment may not have had enough statistical power to bring out a difference, even when one exists. Second, the ERP waveforms may not have been quantified or analyzed in the optimal way. And perhaps most importantly, third, ERPs sample only a subset of the total activity that is going on in the brain at any one time. For its activity to be detectable at the scalp, the elements of a neuronal population must activate (or de-activate) synchronously, and their geometric configuration must be such that their activity summates (that is, they must have an “open field” configuration; see Wood, 1987). Accordingly, the neural activity differentiating the experimental conditions may not have the right dynamic or geometric properties to be detectable on the scalp. For these reasons, when two scalp distributions differ, it is possible to be confident that conditions engaged neurally nonequivalent processes. The converse, however, does not apply.

**Scalp Distribution**
Even when one finds a statistically reliable effect, its interpretation may not be straightforward. Effects on scalp distribution are a good example. As mentioned earlier, reliable differences in scalp distribution allow several possible conclusions. One is that functionally nonequivalent cognitive processes are engaged across conditions or time. Scalp distribution differences can only come about when the patterns of neural activity generating the distributions differ across conditions or time. However, it is unclear from ERP data alone what the exact nature of this difference is.

An ERP effect may be generated by a single, anatomically circumscribed neuronal population, or it may reflect the contribution of multiple, anatomically distributed populations. This means that there is more than one reason why the scalp distributions of two ERP effects may differ. In the simplest case, different distributions may signify the engagement of anatomically distinct generators. Alternatively, scalp distribution effects could reflect differences in the relative contributions of the different components of a common set of generators, in terms either of their strengths or time courses. In the first case, one would conclude that the two effects are truly distinct anatomically. In the second case, the effects might both reflect activity within a
common functional network; whether such a finding constitutes evidence of a strong functional dissociation is arguably less obvious (e.g., Urbach & Kutas, 2002).

**Polarity**
A defining feature of an ERP effect is its polarity. It is important to note, however, that whether an effect is observed to be positive-going or negative-going depends on a variety of nonneurophysiological factors, such as the location of the reference electrode, the baseline against which the effect is compared, and the location and orientation of its intracerebral sources. The polarity of an ERP effect may also vary because of neurophysiological reasons. For example, the orientation of the electromagnetic field generated by the same neuronal population depends both on whether input is inhibitory or excitatory, and whether input is received via synapses distal or proximal to the cell bodies (Wood, 1987). For all of these reasons, in the absence of detailed information about the neural activity underlying it, the polarity of an ERP effect is of no particular neurophysiological or functional significance.

**Intracerebral Sources**
It should be obvious from the foregoing discussions that ERP data recorded from the scalp do not allow direct inferences about either the identity or the spatial location within the brain of the neural activity that gives rise to it. In other words, there is not a transparent relationship between an electrical field observed on the scalp and the brain regions giving rise to that field. For example, if an ERP effect is maximal over frontal scalp sites, this does not necessarily mean that the activity that gives rise to this effect is in frontal cortex. Clearly, it would be of considerable value to be able to discern the intracerebral sources of ERP data. Such knowledge would enhance the functional and neural interpretations of the data, and greatly facilitate its integration with findings from studies using other methods (e.g., fMRI). For detailed discussion of these issues, see chapters 7, 9, and 15 of this volume.

**Amplitude**
In addition to inferences based on scalp distribution, those based on the amplitude of an ERP effect also need qualification. Crucially, amplitude differences can occur in the absence of a change in the strength of the underlying neural activity. ERP waveforms are almost always formed by averaging across multiple EEG epochs, time-locked to a common class of events. Quantification of the waveforms therefore relies on the assumptions underlying the employment of signal averaging. One of these assumptions is that the signal is invariant across epochs. If there is variability in the time of occurrence of the signal (“latency jitter”), and the degree of variability differs between conditions, amplitude differences may occur in the averaged ERP waveforms even though the signal in the constituent epochs does not differ in amplitude. Methods do exist to estimate the signal in single trials rather than the averaged waveform (e.g.,
Childers et al., 1987; chapter 10 of this volume). However, because of the low signal-to-noise ratio, single-trial analyses are only successful when the signal of interest is large. Furthermore, such analyses are inappropriate when the effect of interest is inherent to the difference between different classes of trials (as is the case, for example, for ERP “repetition effects”; e.g., Bentin & Peled, 1990).

Arguably more important is a second way in which the assumption of across-trial invariance can be violated. According to this assumption, differences in amplitude between two ERP effects reflect differences in the degree to which their underlying generators are active, and hence the degree of engagement of the associated cognitive processes (see above). It is possible, however, that such differences merely reflect differences in the proportion of trials carrying an effect of constant amplitude. Under these circumstances, amplitude differences provide information not about the degree to which a process was engaged on any given trial, but about the probability of its engagement. Despite the markedly different theoretical interpretations that follow from these two scenarios, distinguishing between them can be formidably difficult, if not impossible.

**Time Course**

As already discussed, the time at which ERP waveforms diverge provides a measure of the time by which the neural activity, and hence the associated cognitive process, differs between experimental conditions. There are two considerations here. First, the onset of an effect does not necessarily reflect the actual point in time when the brain first distinguishes the conditions. It is possible that neural activity differed before this time, but that the ERPs were not sensitive to this difference (for example, because the activity was not detectable at the scalp). Thus, the onset latency of an ERP effect should be viewed as an upper bound on the time by which cognitive processing started to differ. Second, it is important to note that although we have focused above on onset latency, the full characterization of the time course of an ERP effect may require the estimation of several other parameters as well, for example, latency to peak, rise time, and duration. For some questions, such as when one wants to know not only about when a hypothetical process begins but also how long it lasts, these other parameters may be of equal importance.

**Correlation versus Causation**

All inferences from ERP data, and neuroimaging data in general, are correlational in nature. That is, they allow one to make statements about neural activity that is correlated with some cognitive process, but not whether this activity is necessary for that process to occur. Even when a tight correlation is found between some experimental manipulation and some neural measure, one cannot conclude that the measured activity is a direct manifestation of the cognitive process thought to be associated
with the manipulation. Instead, it may reflect cognitive processes that occur downstream from the process of interest, or be incidental to it.

To determine whether a specific neural correlate is necessary for some cognitive process, one must investigate the consequences of interfering with the correlate. This might be achieved by studying patients with brain lesions, or by temporarily disrupting neural activity with techniques such as Transcranial Magnetic Stimulation (Cowey & Walsh, 2001), or pharmacological manipulations. Although a positive finding does not necessarily mean that the neural activity in question is necessary for the cognitive process of interest (Rugg, 1999), if the process is unaffected by interventions that abolish a neural correlate, one can confidently conclude that the activity in question plays no causal role in the process.

**Interdomain Mapping**

Using any measure of brain activity to understand functional processes requires a conceptualization of how functional states map onto physical brain activity. For example, several of the functional interpretations from ERP data discussed in this chapter are based on the assumption that different patterns of neural activity (as manifested in scalp distribution differences) imply qualitative differences in cognitive processes. Such interpretations are sustainable only when one assumes that there exists a one-to-one mapping between neural activity and cognitive processes. That is, that there is one, and only one, pattern of neural activity that underlies a distinct functional state. If, instead, one assumes that the same functional state can be caused by more than one physical state (as proposed by, for example, Mehler, Morton, & Jusczyk, 1984), it is difficult, if not impossible, to use measures of differential brain activity to infer functional processes (for a related discussion see Price and Friston, 2002; and Friston and Price, 2003).

Even when one assumes one-to-one mapping, different patterns of neural activity may not necessarily reveal distinct cognitive processes. To illustrate this point, Rugg and Coles (1995) described an example from the literature on ERPs and attention. The early deflections in ERP waveforms elicited by visual stimuli are modulated depending on whether the stimuli are attended or ignored. Importantly, the scalp distributions of these modulations differ depending on the visual field in which the stimuli are presented. When a stimulus occurs in the left visual field, the attention-related changes are largest at scalp sites over the right hemisphere, and vice versa (e.g., Mangun & Hillyard, 1995). These qualitative differences in scalp distribution have not, however, led to the conclusion that there exist distinct attentional processes for paying attention to the left versus right visual fields. Instead, the same attentional processes are supported by distinct neuroanatomical pathways (in this case, in homotopic regions of each cerebral hemisphere). Thus, there are at least some circumstances in which distinct patterns of neural activity do not reflect distinct cognitive processes. As a
consequence, the demonstration that experimental conditions are associated with distinct scalp distributions (and therefore distinct patterns of neural activity) is a necessary, but not a sufficient, condition for concluding that distinct cognitive processes are engaged across conditions.

Finally, the question arises as to what constitutes evidence of distinct patterns of neural activity. It is presently unclear just how different two patterns need to be before they should be considered functionally distinct. Given a sufficiently sensitive measure, for example, one could in principle differentiate activity at the level of single neurons. This does not necessarily mean that it would be meaningful to apply a functional interpretation to differences at this level. As things stand at the moment, the trend in the ERP literature is for any statistically significant difference in scalp distribution to be considered of potential functional interest. The spatial resolution with which ERP scalp fields are sampled is continuously increasing, however, and along with it the power to detect subtle differences in scalp distribution. It will be of interest to see whether any limit emerges on the size of an effect that carries functional significance.

Conclusion

ERPs have provided important insights into perceptual, cognitive, and motor functions since the 1960s. In this chapter, we have highlighted the ways in which one can make functional interpretations from ERP data, the assumptions upon which they are based, and some of the difficulties surrounding functional interpretations. Because of their high temporal resolution and low cost, ERPs will likely remain an essential tool in cognitive neuroscience. An important development in future will be the integration of ERP data with neuroimaging tools that allow the specification of where neural activity originates inside the brain.

References


