An Introduction to the Event-Related Potential Technique

Steven J. Luck

The MIT Press

From The MIT Press

MITCogNet

© 2005 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Melior and Helvetica on 3B2 by Asco Typesetters, Hong Kong.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Luck, Stephen J.
An introduction to the event-related potential technique / Stephen J. Luck.
p. cm. — (Cognitive neuroscience)
Includes bibliographical references and index.
ISBN 0-262-12277-4 (alk. paper) — ISBN 0-262-62196-7 (pbk. : alk. paper)
1. Evoked potentials (Electrophysiology) I. Title. II. Series.
QP376.5.L83 2005
616.8′047547—dc22 2005042810

10 9 8 7 6 5 4 3 2 1

2 The Design and Interpretation of ERP Experiments

This chapter discusses some of the central issues in the design and interpretation of ERP experiments. Many such issues are unique to a given research topic (e.g., equating word frequencies in language experiments), but this chapter will focus on a set of principles that are common to most cognitive ERP studies.[1] Throughout the chapter, I will distill the most significant points into a set of rules and strategies for designing and analyzing ERP experiments.

It may seem odd to place a chapter on experimental design and interpretation before the chapters that cover basic issues such as electrodes and averaging. However, the experimental design is the most important element of an ERP experiment, and the principles of experimental design have implications for the more technical aspects of ERP research.

Waveform Peaks versus Latent ERP Components

The term *ERP component* refers to one of the most important and yet most nebulous concepts in ERP research. An ERP waveform unambiguously consists of a series of peaks and troughs, but these voltage deflections reflect the sum of several relatively independent underlying or *latent* components. It is extremely difficult to isolate the latent components so that they can be measured independently, and this is the single biggest roadblock to designing and interpreting ERP experiments. Consequently, one of the keys to successful ERP research is to distinguish between the observable peaks of the waveform and the unobservable latent components. This section describes several of the factors that make it difficult to assess the latent components, along with a set of ''rules'' for avoiding misinterpreting the relationship between the observable peaks and the underlying components.

Voltage Peaks Are not Special

Panels A–C in figure 2.1 illustrate the relationship between the visible ERP peaks and the latent ERP components. Panel A shows an ERP waveform, and panel B shows a set of three latent ERP components that, when summed together, equal the ERP waveform in panel A. When several voltages are present simultaneously in a conductor such as the brain, the combined effect of the individual voltages is exactly equal to their sum, so it is quite reasonable to think about ERP waveforms as an expression of several summed latent components. In most ERP experiments, the researchers want to know how an experimental manipulation influences a specific latent component, but we don't have direct access to the latent components and must therefore make inferences about them from the observed ERP waveforms. This is more difficult than it might seem, and the first step is to realize that the maximum and minimum voltages (i.e., the peaks) in an observed ERP waveform are not necessarily a good reflection of the latent components. For example, the peak latency of peak 1 in the ERP waveform in panel A is much earlier than the peak latency of component C1 in panel B. This leads to our first rule of ERP experimental design and interpretation:

Rule 1. Peaks and components are not the same thing. There is nothing special about the point at which the voltage reaches a local maximum.

In light of this fundamental rule, I am always amazed at how often researchers use peak amplitude and peak latency to measure the magnitude and timing of ERP components. These measures often provide a highly distorted view of the amplitude and timing of the latent components, and better techniques are available for quantifying ERP data (see chapter 5).

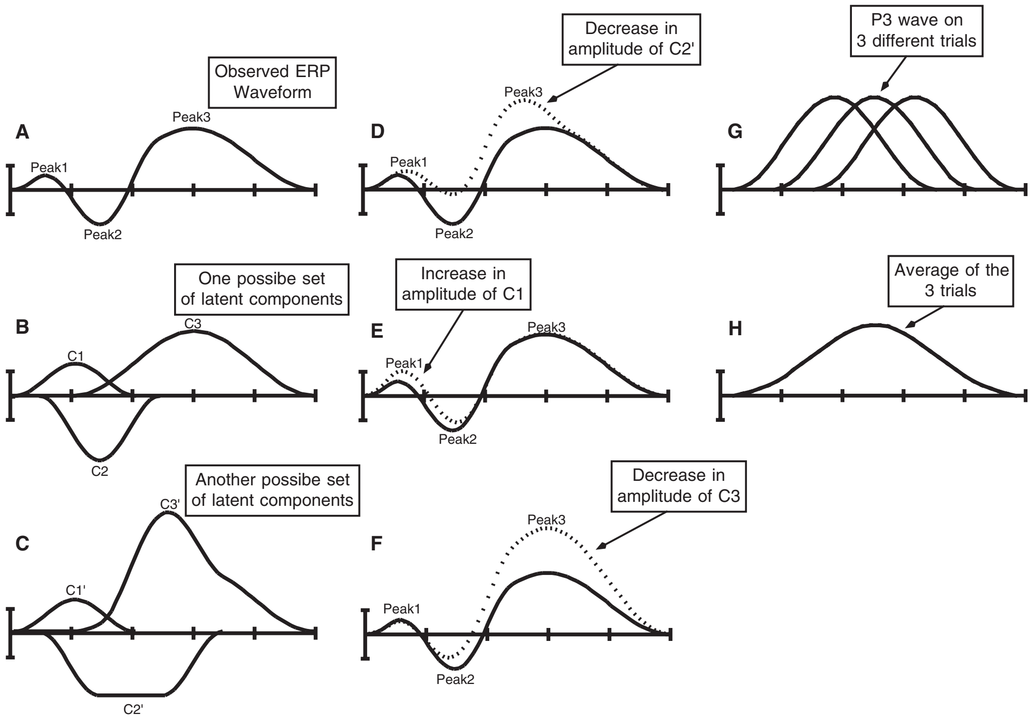

Figure 2.1 Examples of the latent components that may sum together to form an observed ERP waveform. Panels B and C show two different sets of latent components that could underlie the waveform shown in panel A. Panel D shows the effect of decreasing the amplitude of component C2′ by 50 percent (broken line) compared to the original waveform (solid line). Panel E shows how an increase in the amplitude of component C1 (broken line) relative to the original waveform (solid line) can create an apparent shift in the latencies of both peak 1 and peak 2. Panel F shows how an increase in the amplitude of component C3 (broken line) relative to the original waveform (solid line) can influence both the amplitude and the latency of peak 2. Panel G shows a component at three different latencies, representing trial-by-trial variations in timing; panel H shows the average of these three waveforms, which is broader and has a smaller peak amplitude (but the same area amplitude) compared to each of the single-trial waveforms.

Peak Shapes Are not the Same as Component Shapes

Panel C of figure 2.1 shows another set of latent components that also sum together to equal the ERP waveform shown in panel A. In this case, the relatively short duration and hill-like shape of peak 2 in panel A bears little resemblance to the long duration, boxcar-like component C2′ in panel C. This leads to our second rule:

Rule 2. It is impossible to estimate the time course or peak latency of a latent ERP component by looking at a single ERP waveform—there may be no obvious relationship between the shape of a local part of the waveform and the underlying components.

Violations of this rule are especially problematic when comparing two or more ERP waveforms. For example, consider the ERP waveforms in panel D of figure 2.1. The solid waveform represents the sum of the three latent components in panel C (which is the same as the ERP waveform in panel A), and the dashed waveform shows the effects of decreasing component C2′ by 50 percent. To make this a bit more concrete, you can think of these waveforms as the response to an attended stimulus and an unattended stimulus, respectively, such that ignoring the stimulus leads to a 50 percent decline in the amplitude of component C2′. Without knowing the underlying component structure, it would be tempting to conclude from the ERP waveforms shown in panel D that the unattended stimulus elicits not only a decrease in the amplitude of component C2′ but also an increase in the amplitude of component C1′ and a decrease in the latency and an increase in the amplitude of component C3′. In other words, researchers often interpret an effect that overlaps with multiple peaks in the ERP waveform as reflecting changes in multiple underlying components, but this interpretation is often incorrect. Alternatively, you might conclude from the waveforms in panel D that the attentional manipulation adds an additional, long-duration component that would not otherwise be present at all. This would also be an incorrect conclusion, which leads us to:

Rule 3. It is dangerous to compare an experimental effect (i.e., the difference between two ERP waveforms) with the raw ERP waveforms.

Rule 3 applies to comparisons of scalp distribution as well as comparisons of timing. That is, it is not usually appropriate to compare the scalp distribution of an experimental effect with the scalp distribution of the raw ERP waveform. For example, imagine you found that words elicited a more negative response in the N1 latency range than nonwords. To determine whether this effect reflects a change in the amplitude of the exogenous N1 or the addition of a word-specific neural process, it would be tempting to calculate a word-minus-nonword difference wave and compare the scalp distribution of this difference wave (in the N1 latency range) to the scalp distribution of the N1 elicited by nonwords. If the scalp distributions were found to be different, one might conclude that the effect represents the addition of endogenous, word-specific brain activity. But this conclusion would be unwarranted. The nonword-elicited N1 scalp distribution certainly represents overlapping activity from multiple latent components, and a different scalp distribution would be expected for the word-minus-nonword difference wave if a subset of these latent components were larger (or smaller) for words than for nonwords. Moreover, a change in the latency of one of the latent components would also lead to differences in scalp distribution. Thus, if the scalp distribution differs between two conditions, the only certain conclusion is that the experimental manipulation does not simply change the amplitude of all components proportionately.

This raises an important point about the relationship between amplitude and latency. Although the amplitude and latency of a latent component are conceptually independent, amplitude and latency often become confounded when measuring ERP waveforms. This occurs because the latent components overlap in time. Consider, for example, the relatively straightforward correspondence between the peaks in panel A of figure 2.1 and the latent components in panel B of the figure. Panel E of the figure shows the effects of increasing the amplitude of the first latent component on the summed ERP activity. When the amplitude of component A is increased by 50 percent, this creates an increase in the latency of both peak 1 and peak 2 in the summed waveform, and it also causes a decrease in the peak amplitude of peak 2. Panel F illustrates the effect of doubling the amplitude of the component C3, which causes a decrease in the amplitude and the latency of the second peak. Once again, this shows how the peak voltage in a given time range is a poor measure of the underlying ERP components in that latency range. This leads to our next rule:

Rule 4. Differences in peak amplitude do not necessarily correspond with differences in component size, and differences in peak latency do not necessarily correspond with changes in component timing.

Distortions Caused by Averaging

In the vast majority of ERP experiments, the ERP waveforms are isolated from the EEG by means of signal-averaging procedures. It is tempting to think of signal-averaging as a process that simply attenuates the nonspecific EEG, allowing us to see what the single-trial ERP waveforms look like. However, to the extent that the single-trial waveform varies from trial to trial, the averaged ERP may provide a distorted view of the single-trial waveforms, particularly when component latencies vary from trial to trial. Panels G and H of figure 2.1 illustrate this distortion. Panel G illustrates three single-trial ERP waveforms (without any EEG noise), with significant latency differences across trials, and panel H shows the average of those three single-trial waveforms. The averaged waveform differs from the single-trial waveforms in two significant ways. First, it is smaller in peak amplitude. Second, it is more spread out in time. In addition, even though the waveform in panel H is the average of the waveforms in panel G, the onset time of the averaged waveform in panel H reflects the onset time of the earliest single-trial waveform and not the average onset time. This leads to our next rule:

Rule 5. Never assume that an averaged ERP waveform accurately represents the individual waveforms that were averaged together. In particular, the onset and offset times in the averaged waveform will represent the earliest onsets and latest offsets from the individual trials or individual subjects that contribute to the average.

Fortunately, it is often possible to measure ERPs in a way that avoids the distortions created by the signal-averaging process. For example, the area under the curve in the averaged waveform shown in panel H is equal to the average of the area under the single-trial curves in panel G. In most cases, measurements of area amplitude (or mean amplitude) are superior to measurements of peak amplitude. Similarly, it is possible to find the time point that divides the area into two equal halves, and this can be a better measure of latency than peak measures (for more details, see chapter 6).

It is worth mentioning that a very large number of published ERP experiments have violated the five rules presented so far. There is no point in cataloging the cases (especially given that the list would include some of my own papers). However, violations of these rules significantly undermine the strength of the conclusions that one can draw from these experiments. For new students of the ERP technique, it would be worth reading a large set of ERP papers and trying to identify both violations of these rules and methods for avoiding the pitfalls that the rules address.

What Is an ERP Component?

So how can we accurately assess changes in latent components on the basis of the observed ERP waveforms? Ideally, we would like to be able to take an averaged ERP waveform and use some simple mathematical procedure to recover the actual waveforms corresponding to the components that sum together to create the recorded ERP waveform. We could then measure the amplitude and the latency of the isolated components, and changes in one component would not influence our measurement of the other components. Unfortunately, just as there are infinitely many generator configurations that could give rise to a given ERP scalp distribution, there are infinitely many possible sets of latent components that could be summed together to give rise to a given ERP waveform. In fact, this is the basis of Fourier analysis: any waveform can be decomposed into the sum of a set of sine waves. Similarly, techniques such as principal components analysis (PCA) and independent components analysis (ICA) use the correlational structure of a data set to derive a set of basis components that can be added together to create the observed waveforms (for more information, see Donchin & Heffley, 1978; Makeig et al., 1997). Localization techniques can also be used to compute component waveforms at the site of each ERP generator source. Unfortunately, none of these techniques has yet been perfected, as discussed in this section.

All techniques for estimating the latent components are based on assumptions about what a component is. In the early days of ERP research, a component was defined primarily on the basis of its polarity, latency, and general scalp distribution. For example, the P3a and P3b components were differentiated on the basis of the earlier peak latency and more frontal distribution of the P3a component relative to the P3b component. However, polarity, latency, and scalp distribution are superficial features that don't really capture the essence of a component. For example, the peak latency of the P3b component may vary by hundreds of milliseconds depending on the difficulty of the target-nontarget discrimination (Johnson, 1986), and the scalp distribution of the auditory N1 wave depends on the pitch of the eliciting stimulus in a manner that corresponds with the tonotopic map of auditory cortex (Bertrand, Perrin, & Pernier, 1991). Even polarity may vary: the C1 wave, which is generated in primary visual cortex, is negative for upper-field stimuli and positive for lower-field stimuli due to the folding pattern of this cortical area (Clark, Fan, & Hillyard, 1995). Consequently, many investigators now define components in terms of a combination of computational function and neuroanatomical generator site (see, e.g., Näätänen & Picton, 1987). Consistent with this approach, my own definition of the term *ERP component* is:

Scalp-recorded neural activity that is generated in a given neuroanatomical module when a specific computational operation is performed.

By this definition, a component may occur at different times under different conditions, as long as it arises from the same module and represents the same cognitive function (e.g., the encoding of an item into working memory in a given brain area may occur at different delays following the onset of a stimulus because of differences in the amount of time required to identify the stimulus and decide that it is worth storing in working memory). The scalp distribution and polarity of a component may also vary according to this definition, because the same cognitive function may occur in different parts of a cortical module under different conditions (e.g., when a visual stimulus occurs at different locations and therefore stimulates different portions of a topographically mapped area of visual cortex). It is logically possible for two different cortical areas to accomplish exactly the same cognitive process, but this probably occurs only rarely and would lead to a very different pattern of voltages, and so this would not usually be considered a single ERP component.[2]

Techniques such as PCA and ICA use the correlational structure of an ERP data set to define a set of components, and these techniques therefore derive components that are based on functional relationships. Specifically, different time points are grouped together as part of a single component to the extent they tend to vary in a correlated manner, as would be expected for time points that reflect a common cognitive process. The PCA technique, in particular, is problematic because it does not yield a single, unique set of underlying components without additional assumptions (see, e.g., Rosler & Manzey, 1981). That is, PCA really just provides a means of determining the possible set of latent component waveshapes, but additional assumptions are necessary to decide on one set of component waveshapes (and there is typically no way to verify that the assumptions are correct). The ICA technique appears to be a much better approach, because it uses both linear and nonlinear relationships to define the components. However, because ICA is a new technique, it remains to be seen whether it will turn out to be a generally useful means of identifying latent components. In particular, any technique based on identifying correlated versus independent time points will be limited by that fact that when two separate cognitive processes covary, they may be captured as part of a single component even if they occur in very different brain areas and represent different cognitive functions. For example, if all of the target stimuli in a given experimental paradigm are transferred into working memory, an ERP component associated with target detection may always be accompanied by a component associated with working memory encoding, and this may lead ICA to group them together as a single component. Moreover, both PCA and ICA fail when latencies vary across conditions. Thus, correlation-sensitive techniques may sometimes be useful for identifying latent ERP components, but it remains to be seen whether these techniques are generally effective under typical experimental conditions.

Techniques for localizing ERPs can potentially provide measures of the time course of activity within anatomically defined regions. In fact, this aspect of ERP localization techniques might turn out to be just as important as the ability to determine the neuroanatomical locus of an ERP effect. However, as discussed in chapters 1 and 7, there are no general-purpose and foolproof techniques for definitively localizing ERPs at present, and we may never have techniques that allow direct and accurate ERP localization. Thus, this approach to identifying latent ERP components is not generally practical at the present time.

In addition to mathematical techniques for determining the latent components in an ERP waveform, it is also possible to use the world's most powerful pattern analyzer, the human brain. As you gain expertise with specific components and with ERPs in general, you will be able to look at a set of ERP data and make a reasonably good inference about the underlying component structure. This involves considering how the peaks interact with each other across time and across electrode locations and coming up with an interpretation that is consistent with the observed waveforms. Kramer (1985) studied this formally, creating a set of simulated ERP waveforms and asking a set of graduate students and research assistants to make judgments about them. He found that these individuals did a good job of recovering the underlying component structure from the waveforms, but that this depended on the observer's level of experience. Thus, you should become skilled at determining the component structure of ERP waveforms over time. This is an important skill, because you will use your assessment of the latent components to guide your measurement and analysis of the waveforms.

Avoiding Ambiguities in Interpreting ERP Components

The preceding sections of this chapter are rather depressing, because it seems like there is no perfectly general means for measuring latent components from observed ERP waveforms. This is a major problem, because many ERP experiments make predictions about the effects of some experimental manipulation on a given component, and the conclusions of these experiments are valid only if the observed effects really reflect changes in that component. For example, the N400 component is widely regarded as a sensitive index of the degree of mismatch between a word and a previously established semantic context, and it would be nice to use this component to determine which of two sets of words subjects perceive as being more incongruous. If two sets of words elicit different ERP waveforms, it is necessary to know whether this effect reflects a larger N400 for one set or a larger P3 for the other set; otherwise, it is impossible to determine whether the two sets of words differ in terms of semantic mismatch or some other variable (i.e., a variable to which the P3 wave is sensitive). Here I will describe six strategies for minimizing factors that lead to ambiguous relationships between the observed ERP waveforms and the latent components.

Strategy 1. Focus on a Specific Component

The first strategy is to focus a given experiment on only one or perhaps two ERP components, trying to keep as many other components as possible from varying across conditions. If fifteen different components vary, you will have a mess, but variations in a single component are usually tractable. Of course, sometimes a ''fishing expedition'' is necessary when using a new paradigm, but don't count on obtaining easily interpretable results in such cases.

Strategy 2. Use Well-Studied Experimental Manipulations

It is usually helpful to examine a well-characterized ERP component under conditions that are as similar as possible to conditions in which that component has previously been studied. For example, when Marta Kutas first started recording ERPs in language paradigms, she focused on the P3 wave and varied factors such as ''surprise value'' that had previously been shown to influence the P3 wave in predictable ways. Of course, when she used semantic mismatch to elicit surprise, she didn't observe the expected P3 wave but instead discovered the N400 component. However, the fact that her experiments were so closely related to previous P3 experiments made it easy to determine that the effect she observed was a new negative-going component and not a reduction in the amplitude of the P3 wave.

Strategy 3. Focus on Large Components

When possible, it is helpful to study large components such as P3 and N400. When the component of interest is very large compared to the other components, it will dominate the observed ERP waveform, and measurements of the corresponding peak in the ERP waveform will be relatively insensitive to distortions from the other components.

Strategy 4. Isolate Components with Difference Waves

It is often possible to isolate the component of interest by creating difference waves. For example, imagine that you are interested in assessing the N400 for two different noun types, count nouns (e.g., *cup*) and mass nouns (e.g., *water*). The simple approach to this might be to present one word per second, with count nouns and mass nouns randomly intermixed. This would yield two ERP waveforms, one for count nouns and one for mass nouns, but it would be difficult to know if any differences observed between the count noun and mass noun waveforms were due to difference in N400 amplitude or due to differences in some other ERP component.

To isolate the N400, you could redesign the experiment so that each trial contained a sequence of two words, a context word and a target word, with count noun target word on some trials and a mass noun target word on others. In addition, the context and target words would sometimes be semantically related and sometimes be semantically unrelated. You would then have four types of trial types:

Count noun, related to context word (e.g., ‘‘plate . . . cup’’)
Mass noun, related to context word (e.g., ‘‘rain . . . water’’)
Count noun, unrelated to context word (e.g., ‘‘sock . . . cup’’)
Mass noun, unrelated to context word (e.g., ‘‘garbage . . . water’’)

You could then isolate the N400 by constructing difference waves in which the ERP waveform elicited by a given word when it was preceded by a semantically related context word is subtracted from the ERP waveform elicited by that same word when preceded by a semantically unrelated context word. Separate difference waves would be constructed for count nouns and mass nouns (unrelated minus related count nouns and unrelated minus related mass nouns). Each of these difference waves should be dominated by a large N400 component, with little or no contribution from other components (because most other components aren't sensitive to semantic mismatch). You could then see if the N400 was larger in the count noun difference wave or in the mass noun difference wave (I describe a real application of this general approach at the very end of this chapter—see Vogel, Luck, & Shapiro, 1998 for details).

Although this approach is quite powerful, it has some limitations. First, difference waves constructed in this manner may contain more than one ERP component. For example, there may be more than one ERP component that is sensitive to the degree of semantic mismatch, so an unrelated-minus-related difference wave might consist of two or three components rather than just one. However, this is still a vast improvement over the raw ERP waveforms, which probably contain at least ten different components. The second limitation of this approach is that it is sensitive to interactions between the variable of interest (e.g., count nouns versus mass nouns) and the factor that is varied to create the difference waves (e.g., semantically related versus unrelated word pairs). If, for example, the N400 amplitude is 1 μV larger for count nouns than for mass nouns, regardless of the degree of semantic mismatch, then count noun difference waves will be identical to the mass noun difference waves. Fortunately, when two factors influence the same ERP component, they are likely to interact multiplicatively. For example, N400 amplitude might be 20 percent greater for count nouns than for mass nouns, leading to a larger absolute difference in N400 amplitude when the words are unrelated to the context word than when they are related. Of course, the interactions could take a more complex form that would lead to unexpected results. For example, count nouns could elicit a larger N400 than mass nouns when the words are unrelated to the context word, but they might elicit a smaller N400 when the words are related to the context word. Thus, although difference waves can be very helpful in isolating specific ERP components, care is necessary when interpreting the results.

I should also mention that the signal-to-noise ratio of a difference wave will be lower than those of the original ERP waveforms. Specifically, if the original waveforms have similar noise levels, then the noise in the difference wave will be larger by a factor of the square root of two (i.e., approximately 40 percent larger).

Strategy 5. Focus on Components That Are Easily Isolated

The previous strategy advocated using difference waves to isolate ERP components, and one can refine this by focusing on certain ERP components that are relatively easy to isolate. The best example of this is the lateralized readiness potential (LRP), which reflects movement preparation and is distinguished by its contralateral scalp distribution. Specifically, the LRP in a given hemisphere is more negative when a movement of the contralateral hand is being prepared than when a movement of the ipsilateral hand is being prepared, even if the movements are not executed. In an appropriately designed experiment, only the motor preparation will lead to lateralized ERP components, making it possible to form difference waves in which all ERPs are subtracted away except for those related to lateralized motor preparation (see Coles, 1989; Coles et al., 1995). Similarly, the N2pc component for a given hemisphere is more negative when attention is directed to the contralateral visual field than when it is directed to the ipsilateral field, even when the evoking stimulus is bilateral. Because most of the sensory and cognitive components are not lateralized in this manner, the N2pc can be readily isolated (see, e.g., Luck et al., 1997; Woodman & Luck, 2003).

Strategy 6. Component-Independent Experimental Designs

The best strategy is to design experiments in such a manner that it does not matter which latent ERP component is responsible for the observed changes in the ERP waveforms. For example, Thorpe and colleagues (1996) conducted an experiment in which they asked how quickly the visual system can differentiate between different classes of objects. To answer this question, they presented subjects with two classes of photographs, pictures that contained animals and pictures that did not. They found that the ERPs elicited by these two classes of pictures were identical until approximately 150 ms, at which point the waveforms diverged. From this experiment, it is possible to infer that the brain can detect the presence of an animal in a picture by 150 ms, at least for a subset of pictures (note that the onset latency represents the trials and subjects with the earliest onsets and not necessarily the average onset time). This experimental effect occurred in the time range of the N1 component, but it may or may not have been a modulation of that component. Importantly, the conclusions of this study do not depend at all on which latent component the experimental manipulation influenced. Unfortunately, it is rather unusual to be able to answer a significant question in cognitive neuroscience using ERPs in a component-independent manner, but one should use this approach whenever possible. I will provide additional examples later in this chapter.

Avoiding Confounds and Misinterpretations

The problem of assessing latent components on the basis of observed ERP waveforms is usually the most difficult aspect of the design and interpretation of ERP experiments, and this problem is particularly significant in ERP experiments. There are other significant experimental design issues that are applicable to a wide spectrum of techniques, but are particularly salient in ERP experiments; these will be the focus of this section.

The most fundamental principle of experimentation is to make sure that a given experimental effect has only a single possible cause. One part of this principle is to avoid confounds, but a subtler part is to make sure that the experimental manipulation doesn't have secondary effects that are ultimately responsible for the effect of interest. For example, imagine that you observe that the mass of a beaker of hot water is less than the mass of a beaker of cold water. This might lead to the erroneous conclusion that hot water has a lower mass than cool water, even though the actual explanation is that some of the heated water turned to steam, which escaped through the top of the beaker. To reach the correct conclusion, it is necessary to seal the beakers so that water does not escape. Similarly, it is important to ensure that experimental manipulations in ERP experiments do not have unintended side effects that lead to an incorrect conclusion.

To explore the most common problems that occur in ERP experiments, let's consider a thought experiment that examines the effects of stimulus discriminability on P3 amplitude. In this experiment, letters of the alphabet are presented foveally at a rate of one per second and the subject is required to press a button whenever the letter Q is presented. A Q is presented on 10 percent of trials and a randomly selected non-Q letter is presented on the other 90 percent. In addition, the letter Q never occurs twice in succession. In one set of trial blocks, the stimuli are bright and therefore easy to discriminate (the bright condition), and in another set of trial blocks the stimuli are very dim and therefore difficult to discriminate (the dim condition).

There are several potential problems with this seemingly straightforward experimental design, mainly due to the fact that the target letter (Q) differs from the nontarget letters in several ways. First, the target category occurs on 10 percent of trials whereas the nontarget category occurs on 90 percent of trials. This is one of the two intended experimental manipulations (the other being target discriminability). Second, the target and nontarget letters are different from each other. Not only is the target letter a different shape from the nontarget letters—and might therefore elicit a somewhat different ERP waveform—the target letter also occurs more frequently than any of the individual nontarget letters. To the extent that the visual system exhibits long-lasting and shape-specific adaptation to repeated stimuli, it is possible that the response to the letter Q will become smaller than the response to the other letters. These physical stimulus differences probably won't have a significant effect on the P3 component, but they could potentially have a substantial effect on earlier components (for a detailed example, see experiment 4 of Luck & Hillyard, 1994a).

An important principle of ERP experimental design that I learned from Steve Hillyard is that you should be very careful to avoid confounding differences in psychological factors with subtle differences in stimulus factors. In this example experiment, for example, the probability difference between the target and nontarget letters—which is intended to be a psychological manipulation—is confounded with shape differences and differences in sensory adaptation. Although such confounds primarily influence sensory responses, these sensory responses can last for hundreds of milliseconds, and you will always have the nagging suspicion that your P3 or N400 effects reflect a sensory confound rather than your intended psychological manipulation. The solution to this confound is to make sure that your manipulations of psychological variables are not accompanied by any changes in the stimuli (including the sequential context of the stimulus of interest). This is typically accomplished by using the same stimulus sequences in two or more conditions and using verbal or written instructions to manipulate the subject's task. A statement that is prominently displayed on a wall in my laboratory summarizes this:

The Hillyard Principle—Always compare ERPs elicited by the same physical stimuli, varying only the psychological conditions.

Of course, implementing the Hillyard Principle is not always possible. For example, you may want to examine the ERPs elicited by closed-class words (articles, prepositions, etc.) and open-class words (nouns, verbs, etc.); these are by definition different stimuli. But you should be very careful that any ERP differences reflect the psychological differences between the stimuli and not low-level physical differences (e.g., word length). I can tell you from experience that every time I have violated the Hillyard Principle, I have later regretted it and ended up running a new experiment. Box 2.1 provides two examples of confounds that somehow crept into my own experiments.

A third difference between the target and nontarget letters is that subjects make a response to the targets and not to the nontargets. Consequently, any ERP differences between the targets and nontargets could be contaminated by motor-related ERP activity. A fourth difference between the targets and the nontargets is that because the target letter never occurred twice in succession, the target letter was always preceded by a nontarget letter, whereas nontarget letters could be preceded by either targets or nontargets. This is a common practice, because the P3 to the second of two targets tends to be reduced in amplitude. This is usually a bad idea, however, because the response to a target is commonly very long-lasting and extends past the next stimulus and therefore influences the waveform recorded for the next stimulus. Thus, there may appear to be differences between the target and nontarget waveforms in the N1 or P2 latency ranges that actually reflect the offset of the P3 from the previous trial, which is present only in the nontarget waveforms under these conditions. This type of differential overlap occurs in many ERP experiments, and it can be rather subtle. (For an extensive discussion of this issue, see Woldorff, 1988.)

A fifth difference between the targets and the nontargets arises when the data are averaged and a peak amplitude measure is used to assess the size of the P3 wave. Specifically, because there are many more nontarget trials than target trials, the signal-to-noise ratio is much better for the nontarget waveforms. The maximum amplitude of a noisy waveform will tend to be greater than the

Box 2.1 Examples of Subtle Confounds

It may seem simple to avoid physical stimulus confounds, but these confounds are often subtle. Let me give two examples where I made mistakes of this nature. Many years ago, I conducted a series of experiments in which I examined the ERPs elicited by visual search arrays consisting of seven randomly positioned ''distractor'' bars of one orientation and one randomly positioned ''pop-out'' bar of a different orientation. In several experiments, I noticed that the P1 wave tended to be slightly larger over the hemisphere contralateral to the pop-out item relative to the ipsilateral hemisphere. I thought this might reflect an automatic capture of attention by the pop-out item, although this didn't fit very well with what we knew about the time course of attention. Fortunately, Marty Woldorff suggested that this effect might actually reflect a physical stimulus confound. Specifically, the location of the pop-out bar on one trial typically contained an opposite-orientation distractor bar on the previous trial, whereas the location of a distractor bar on one trial typically contained a same-orientation distractor bar on the previous trial. Thus, the response to the pop-out bar might have been larger than the response to the distractor bars because the neurons coding the pop-out bar wouldn't have just responded to a stimulus of the same orientation, and this led to a larger response contralateral to the pop-out bar because these responses are usually larger at contralateral than at ipsilateral sites. At first, this seemed unlikely because the screen was blank for an average of 750 ms between trials. However, when I tested Marty's hypothesis with an additional experiment, it turned out that he was correct (see experiment 4 of Luck & Hillyard, 1994a).

I encountered another subtle physical stimulus confound several years later when Massimo Girelli and I were examining whether the N2pc component in a visual search task differed for pop-out stimuli in the upper visual field compared to the lower visual field. Because this question directly involves a physical stimulus manipulation, we knew we had to use a trick to avoid direct physical stimulus confounds. The trick was to use two differently colored pop-out items on each trial, one in the left visual field and one in the right visual field, and to ask the subjects to attend to one color (e.g., the red pop-out) on some trial blocks and to attend to the other color (e.g., the green pop-out) on others. Because the N2pc component is lateralized with respect to the direction of attention, we could then make a difference wave in which the physical stimulus was held constant (e.g., red pop-out on the left and green pop-out on the right) and attention was varied (e.g., attend-red versus attend-green). We expected that all of the purely sensory responses would be subtracted away in this difference wave because the physical stimulus was identical and only attention was varied. This is a useful trick that we have

Box 2.1 (continued)

used several times, but in this case we failed to implement it correctly. Specifically, the goal of our experiment was to compare the responses to upper versus lower field stimuli, and we therefore subdivided our trials as a function of the vertical position of the target item. Unfortunately, when the target was in the upper field, the nontarget pop-out item could be either in the upper field or the lower field. We had coded the vertical position of the attended pop-out item but not the vertical position of the unattended pop-out item, and it was therefore impossible to perform subtractions on exactly the same physical stimuli (e.g., red in upper-left and green in upper-right when red was attended versus when green was attended). When we performed subtractions that were not properly controlled, we obtained a pattern of results that was simply weird, and we had to run the experiment a second time with the appropriate codes.

maximum amplitude of a clean waveform due purely to probability, and a larger peak amplitude for the target waveform could therefore be caused solely by its poorer signal-to-noise ratio even if the targets and nontargets elicited equally large responses.

The manipulation of stimulus brightness is also problematic, because this will influence several factors in addition to stimulus discriminability. First, the brighter stimuli are, well, brighter than the dim stimuli, and this may create differences in the early components that are not directly related to stimulus discriminability. Second, the task will be more difficult with the dim stimuli than with the bright stimuli. This may induce a greater state of arousal during the dim blocks than during the bright blocks, and it may also induce strategy differences that lead to a completely different set of ERP components in the two conditions. A third and related problem is that reaction times will be longer in the dim condition than in the bright condition, and any differences in the ERP waveforms between these two conditions could be due to differences in the time course of motor-related ERP activity (which overlaps with the P3 wave).

There are two main ways to overcome problems such as these. First, you can avoid many of these problems by designing the experiment differently. Second, it is often possible to demonstrate that a potential confound is not actually responsible for the experimental effect; this may involve additional analyses of the data or additional experiments. As an illustration, let us consider several steps that you could take to address the potential problems in P3 experiment described above:

1. You could use a different letter as the target for each trial block, so that across the entire set of subjects, all letters are approximately equally likely to occur as targets or nontargets. This solves the problem of having different target and nontarget shapes.
2. To avoid differential visual adaptation to the target and nontarget letters, you could use a set of ten equiprobable letters, with one serving as the target and the other nine serving as nontargets. Each letter would therefore appear on 10 percent of trials. If it is absolutely necessary that one physical stimulus occurs more frequently than another, it is possible to conduct a sequential analysis of the data to demonstrate that differential adaptation was not present. Specifically, trials on which a nontarget was preceded by a target can be compared with trials on which a nontarget was preceded by a nontarget. If no difference is obtained—or if any observed differences are unlike the main experimental effect—then the effects of stimulus probability are probably negligible.
3. Rather than asking the subjects to respond only to the targets, instruct them to make one response for targets and another for nontargets. Target and nontarget RTs are likely to be different, so some differential motor activity may still be present for targets versus nontargets, but this is still far better than having subjects respond to the targets and not to the nontargets.
4. It would be a simple matter to eliminate the restriction that two targets cannot occur in immediate succession, thus avoiding the possibility of differential overlap from the preceding trial. However, if it is necessary to avoid repeating the targets, it is possible to con-

struct an average of the nontargets that excludes trials preceded by a target. If this is done, then both the target and the nontarget waveforms will contain only trials on which the preceding trial was a nontarget (as in step 2).

5. There are two good ways to avoid the problem that peak amplitudes tend to be larger when the signal-to-noise ratio is lower. First, as discussed above, the peak of an ERP waveform bears no special relationship to the corresponding latent component, so this problem can be solved by measuring the mean amplitude over a predefined latency range rather than the peak amplitude. As discussed in chapter 4, mean amplitude has several advantages over peak amplitude, and one of them is that the measured amplitude is not biased by the number of trials. Of course, mean amplitude measures exhibit increased variance as the signal-to-noise ratio decreases, but the expected value does not vary. If, for some reason, it is necessary to measure peak amplitude rather than mean amplitude, it is possible to avoid biased amplitude measures by creating the nontarget average from a randomly selected subset of the nontarget trials such that the target and nontarget waveforms reflect the same number of trials.
6. There is no simple way to compare the P3 elicited by bright stimuli versus dim stimuli without contributions from simple sensory differences. However, simple contributions can be ruled out by a control experiment in which the same stimuli are used but are viewed during a task that is unlikely to elicit a P3 wave (e.g., counting the total number of stimuli, regardless of the target-nontarget category). If the ERP waveforms for the bright and dim stimuli in this condition differ only in the 50–250 ms latency range, then the P3 differences observed from 300–600 ms in the main experiment cannot easily be explained by simple sensory effects and must instead reflect an interaction between sensory factors (e.g., perceptibility) and cognitive factors (e.g., whatever is responsible for determining P3 amplitude).
7. The experiment should also be changed so that the bright and dim stimuli are randomly intermixed within trial blocks. In this way,

the subject's state at stimulus onset will be exactly the same for the easy and difficult stimuli. This also tends to reduce the use of different strategies.

8. It is possible to use additional data analyses to test whether the different waveforms observed for the dim and bright conditions are due to differences in the timing of the concomitant motor potentials (which is plausible whenever RTs differ between two conditions). Specifically, if the trials are subdivided into those with fast RTs and those with slow RTs, it is possible to assess the size and scalp distribution of the motor potentials. If the difference between trials with fast and slow RTs is small compared to the main experimental effect, or if the scalp distribution of the difference is different from the scalp distribution of the main experimental effect, then this effect probably cannot be explained by differential motor potentials.

Most of these strategies are applicable in many experimental contexts, and they reflect a set of general principles that are very widely applicable. I will summarize these general principles in some additional rules:

Rule 6. Whenever possible, avoid physical stimulus confounds by using the same physical stimuli across different psychological conditions (the Hillyard Principle). This includes ''context'' confounds, such as differences in sequential order.

Rule 7. When physical stimulus confounds cannot be avoided, conduct control experiments to assess their plausibility. Never assume that a small physical stimulus difference cannot explain an ERP effect.

Rule 8. Be cautious when comparing averaged ERPs that are based on different numbers of trials.

Rule 9. Be cautious when the presence or timing of motor responses differs between conditions.

Rule 10. Whenever possible, experimental conditions should be varied within trial blocks rather than between trial blocks.

Examples from the Literature

In this section, I will discuss three experiments that used ERPs to answer a significant question in cognitive neuroscience. In each case, the experiments were designed in such a manner that the conclusions did not depend on identifying specific ERP components, making the results much stronger than is otherwise possible.

Example 1: Auditory Selective Attention

Background Given that I did my graduate work with Steve Hillyard, I was imprinted at an early age on the experimental design that he and his colleagues developed in 1973 to address the classic ''locus-of-selection'' question. This question asks whether attention operates at an early stage of processing, allowing only selected inputs to be perceived and recognized (the *early selection* position of investigators such as Broadbent, 1958; Treisman, 1969), or whether attention instead operates at a late stage such that all incoming sensory events receive equal perceptual processing but only selected stimuli reach decision processes, memory, and behavioral output systems (the *late selection* position of investigators such as Deutsch & Deutsch, 1963; Norman, 1968).

Two factors have made this a difficult question to address with traditional behavioral measures. First, it is difficult to assess the processing of an ignored stimulus without asking the subjects to respond to it, in which case the stimulus may become attended rather than ignored. Second, if responses to ignored stimuli are slower or less accurate than responses to attended stimuli, it is difficult to determine whether this reflects an impairment in sensory processes or an impairment of higher level decision, memory, or response processes. ERPs, in contrast, are particularly well suited for solving both of these problems. First, ERPs are easily recorded in the absence of an overt response, making them ideal for monitoring the processing of an ignored stimulus. Second, ERPs provide precise information about the time course of processing, making them ideal for answering questions about the stage of processing at which an experimental effect occurs. In the case of attention, for example, it is possible to determine whether the early sensory-evoked ERP components are suppressed for ignored stimuli relative to attended stimuli, which would be consistent with the early selection hypothesis, or whether the early ERP components are identical for attended and ignored stimuli and only the later ERP components are sensitive to attention.

Experimental Design Several studies in the 1960s and early 1970s used this logic to address the locus-of-selection question, but these experiments had various methodological shortcomings that made them difficult to interpret. Hillyard, Hink, Schwent, and Picton (1973) reported an experiment that solved these problems and provided unambiguous evidence for early selection. Figure 2.2A illustrates this experiment. Subjects were instructed to attend to the left ear in some trial blocks and the right ear in others. A rapid sequence of tone pips was then presented, with half of the stimuli presented in each ear. To make the discrimination between the two ears very easy, the tones were presented at a different pitch in each ear. Subjects were instructed to monitor the attended ear and press a button whenever a slightly higher pitched ''deviant'' target tone was detected in that ear, which occurred infrequently and unpredictably. Higher pitched tones were also presented occasionally in the ignored ear, but subjects were instructed not to respond to these ''ignored deviants.''

In some prior ERP experiments, subjects were required to respond to or count all attended stimuli and withhold responses to all ignored stimuli (including internal responses such as counting). As a result, any differences in the ERPs evoked by attended and ignored stimuli could have been due to response-related activity that was present for the attended ERPs but absent for ignored ERPs (see rule 9). Hillyard et al. (1973) avoided this problem by presenting both target and nontarget stimuli in both the attended and ignored ears and asking the subjects to count the targets in

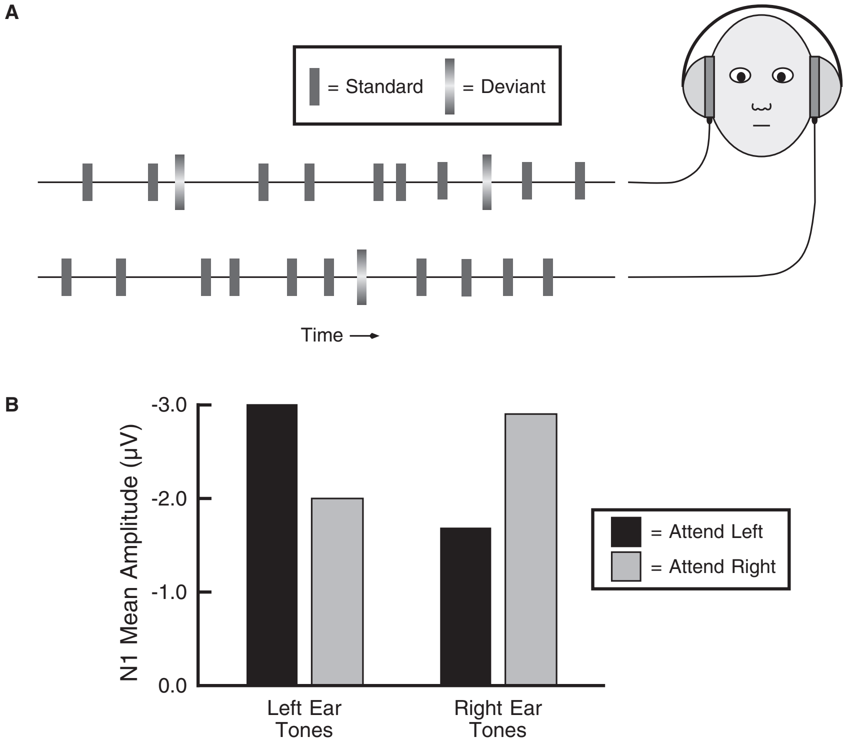

Figure 2.2 Experimental paradigm (A) and results (B) from the study of Hillyard et al. (1973). Subjects listened to streams of tone pips in the left ear and right ear. Most of the tones were a standard frequency (800 Hz in the left ear and 1500 Hz in the right ear), but occasional deviant tones were presented at a slightly higher frequency (840 Hz left; 1560 Hz right). Subjects were instructed to attend to the left ear for some trial blocks and to the right ear for others, and counted the number of deviant tones in the attended ear. The average N1 amplitude was measured for the standard tones and used as an index of sensory processing. Left ear tones elicited a larger N1 wave when attention was directed to the left ear than when attention was directed to the right ear; conversely, right ear tones elicited a larger N1 wave when attention was directed to the right ear than when attention was directed to the left ear. (Reproduced by permission from Luck, 1998a. © 1998 Psychology Press.)

the attended ear. The analyses were focused on the nontargets, to which subjects made neither an external response such as a button press nor an internal response such as counting. To ensure that subjects attended to the nontargets in the attended ear, even though no response was required for them, targets and nontargets within an ear were difficult to discriminate from each other (but easy to discriminate from the stimuli in the other ear). Thus, subjects were motivated to focus attention onto all of the stimuli presented in one ear and ignore all stimuli within the other ear. The main experimental question was whether the early sensory ERP components evoked by a nontarget stimulus presented in the attended ear would be larger than those evoked by a nontarget stimulus presented in the ignored ear.

The sensory ERP components are highly sensitive to the physical characteristics of the evoking stimulus. As a result, one cannot legitimately compare the ERP evoked by an attended tone in the left ear with an ignored tone in the right ear: any differences between these ERPs could be due to differences between the two ears that have nothing to do with attention (see rule 6). The design Hillyard et al. (1973) employed circumvents this problem by allowing the ERP elicited by the same physical stimulus to be compared under different psychological conditions. For example, one can compare the ERP evoked by a left nontarget during attend-left blocks with the ERP evoked by the same left nontarget during attend-right blocks. Because both cases use the same stimulus, any differences in the ERPs between the attend-left and attend-right conditions must be due to differences in attentional processing.

In some attention experiments, the investigators compare an ‘‘active’’ condition, in which the subject responds to the stimuli, with a ‘‘passive’’ condition, in which the subject makes no response to the stimuli and perhaps engages in a distracting activity such as reading a book. Frequently, however, the task in the active condition is much more demanding than the distraction task in the passive condition, leading to greater overall arousal during the active condition. If we compare the ERPs in the two conditions, any differences might be due to these global arousal differences rather than selective changes in stimulus processing. This is a violation of rule 10. To ensure that differences in global arousal would not interfere with their study, Hillyard et al. (1973) compared ERPs evoked during equally difficult attend-left and attend-right conditions rather than active and passive conditions.

Results Now that we have discussed the logic behind this study, let us consider the results. As figure 2.2B shows, the N1 component was found to be larger for attended stimuli than for ignored stimuli.[3] Specifically, the N1 elicited by left ear tones was larger when the left ear was attended than when the right ear was attended, and the N1 elicited by right ear tones was larger when the right ear was attended than when the left ear was attended. These effects began approximately 60–70 ms after stimulus onset and peaked at approximately 100 ms poststimulus. In this manner, Hillyard et al. (1973) were able to demonstrate that, at least under certain conditions, attention can influence the processing of a stimulus within the first 100 milliseconds after stimulus onset, which is consistent with the early selection hypothesis.

Larger Issues I would like to emphasize three aspects of this study. First, it was specifically designed to address an existing and significant question that had previously eluded investigators. This contrasts with many ERP experiments in which it seems as if the authors simply took an interesting cognitive paradigm and ran it while recording ERPs to ''see what happens.'' This approach, although sometimes a useful first step, rarely leads to important conclusions about cognitive or neural issues. Second, this study does not rely on identifying specific ERP components (see strategy 6). The ERPs elicited by the attended and ignored stimuli diverged around 60 to 70 ms poststimulus, and it is this timing information that is the crucial result of the study, and not the fact that the effect occurred in the latency range of the N1 component. In fact, there has been much controversy about whether attention influences the

N1 component per se, but the finding of an effect within 100 ms of stimulus onset will continue to be important regardless of the outcome of this dispute.

A third important aspect of this study is that it used ERPs to assess the processing of stimuli for which subjects made no overt response. This is one of the main advantages of the ERP technique over behavioral measures, and many of the most significant ERP experiments have exploited this ability of ERPs to be used for the ''covert monitoring'' of cognitive processes.

Example 2: Partial Information Transmission

Background Most early models of cognition assumed that simple cognitive tasks were accomplished by means of a sequence of relatively independent processing stages. These models were challenged in the late 1970s by models in which different processes could occur in parallel, such as McClelland's (1979) cascade model and Eriksen and Schultz's (1979) continuous flow model. At the most fundamental level, these new models differed from the traditional models in that they postulated that partial results from one process could be transmitted to another process, such that the second process could begin working before the first process was complete. Traditional discrete-stage models, in contrast, assumed that a process worked until it achieved its result, which was then passed on to the next stage. This is a crucial distinction, but it is very difficult to test without a direct means of observing the processes that are thought to overlap. This is exactly the sort of issue that ERPs can easily address.

By coincidence, two different laboratories conducted similar ERP studies of this issue at about the same time, and both used the lateralized readiness potential (LRP) as a means of assessing whether response systems can become activated before stimulus identification is complete (Miller & Hackley, 1992; Osman et al., 1992). Both were excellent studies, but here I will focus on the one conducted by Miller and Hackley. These investigators tested the specific hypothesis that subjects will begin to prepare a response to a stimulus based on the most salient aspects of the stimulus, even if they later withheld that response because of additional information extracted from the stimulus. In other words, they predicted that motor processing may sometimes begin before perceptual processing is complete, which would be incompatible with traditional discrete-stage models of cognition.

Experimental Design To test this hypothesis, they presented subjects with one of four stimuli on each trial: a large S; a small S; a large T; or a small T. Subjects responded with one hand for S and with the other hand for T, but they responded only to one of the two sizes; they gave no response for the other size (half of the subjects responded to large stimuli and half to small). Thus, this paradigm was a hybrid of a go/no-go design (go for one size and no-go for the other) and a two-alternative forced choice design (one response for S, a different response for T). The shape difference was very salient, but the size difference was relatively difficult to discriminate. Consequently, subjects could begin to prepare a given hand to respond as soon as they discriminated the shape of the letter, and they could later choose to emit or withhold this response when they eventually discriminated the size of the letter.

To determine if the subjects prepared a specific hand for response on the basis of letter shape, even on trials when they made no response because the letter was the wrong size, Miller and Hackley used the LRP component. The LRP component, which reflects response preparation, is particularly useful for this purpose because it is lateralized with respect to the responding hand. Miller and Hackley's (1992) paper provides a wonderful review of the LRP literature, and I highly recommend reading it. Here I will mention two essential aspects of the LRP that were essential for testing their hypothesis. First, the LRP begins before muscle contractions begin and can occur in the absence of an overt response, indicating that it reflects response preparation. Second, the LRP is larger over the hemisphere contralateral to the movement being prepared. Consequently, the presence of an LRP is virtually absolute proof that the brain has begun to differentially prepare one hand for responding. In other words, there is no way to get a consistently larger response over the hemisphere contralateral to a given hand unless the brain has begun to prepare a response for that hand. Thus, Miller and Hackley predicted that an LRP would be observed contralateral to the hand indicated by the shape of the letter, even if the size of the letter indicated that no response should be made (note that EMG recordings were used to ensure that absolutely no response was made on these trials). When the size of the letter was finally perceived, the LRP would then be terminated if it indicated that no response should be made.

Before I describe the results, I want to mention a technique to isolate the LRP component and to ensure that it purely reflects contralateral hand preparation. Imagine that we compare the ERPs recorded at left and right premotor electrode sites just prior to a left-hand response. If the voltage were more negative over the right hemisphere than the left hemisphere, we might conclude that the brain was preparing the left-hand response, leading to a greater negativity over the right hemisphere. This conclusion would be unwarranted, however, because the same result would be obtained if the right hemisphere yielded a more negative response regardless of which hand was being prepared. Similarly, imagine that we compare the response to left-hand and right-hand responses at a left premotor electrode site. If the voltage were more negative prior to left-hand responses than prior to right-hand responses, we might again conclude the brain was preparing the left-hand response. This would again be unwarranted, because the same result would be obtained if left-hand responses yielded more negative voltages than right-hand responses over both hemispheres.

To isolate activity that purely reflects lateralized response preparation, it is necessary to record waveforms corresponding to the different combinations of left and right hemisphere electrode (E_{left} and E_{right}) and left and right hand response (R_{left} and R_{right}). There are four combinations: left hemisphere electrode with left hand response ($E_{left}R_{left}$), left hemisphere electrode with right hand response ($E_{left}R_{right}$), right hemisphere electrode with left hand response ($E_{right}R_{left}$), and right hemisphere electrode with right hand response ($E_{right}R_{right}$). As discussed by Coles (1989), the LRP can then be isolated by creating a difference wave according to this formula:

$$\mathrm{LRP} = [(E_{right}R_{left} - E_{left}R_{left}) + (E_{left}R_{right} - E_{right}R_{right})] \div 2 \qquad (2.1)$$

This formula computes the average of the contralateral-minus-ipsilateral difference for left-hand and right-hand responses, eliminating any overall differences between the left and right hemispheres (independent of hand) and any overall differences between the left and right hands (independent of hemisphere). All that remains is the extent to which the response is generally larger over the hemisphere contralateral to the hand being prepared, regardless of which hand is being prepared or which hemisphere is being recorded. This procedure effectively isolates the LRP from the vast majority of ERP components, which do not show this intricate interaction between hemisphere and hand (see strategy 5).

Results Figure 2.3 shows the results of this study. Panel A shows the averaged ERP waveforms from the left- and right-hemisphere electrodes sites for left- and right-hand responses when the stimulus was the appropriate size for a response (“Go Trials”). The electrode sites used here are labeled C3′ and C4′ to indicate premotor sites just lateral to the standard C3 and C4 sites in the left and right hemispheres, respectively. The ERP waveform at the left-hemisphere site was more negative on trials with a right-hand response than on trials with a left-hand response, and the complementary pattern was observed at the right-hemisphere site. This effect began approximately 200 ms poststimulus and continued for at least 800 ms, which is typical for the LRP component.

Panel B of figure 2.3 shows the responses observed when the size of the stimulus indicated that no response should be made (“No-Go

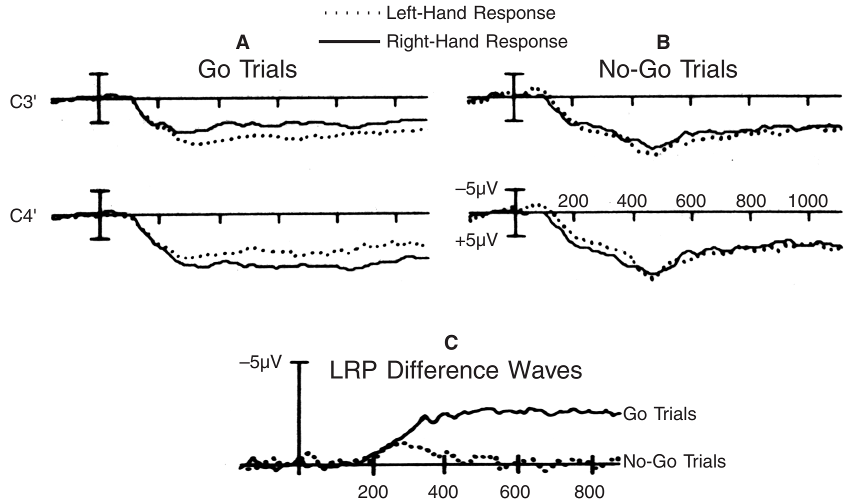

Figure 2.3 Grand-average ERP waveforms from the study of Miller and Hackley (1992). The waveforms in panels A and B reflect trials on which the size of the letter indicated that a response should be made ("Go Trials," panel A) or should not be made ("No-Go Trials," panel B). These waveforms were recorded from left central (C3′) and right central (C4′) electrode sites. Broken lines reflect trials on which the shape of the letter indicated a left-hand response, and solid lines reflect trials on which the shape of the letter indicated a right-hand response. Panel C shows LRP difference waves for the go and no-go trials. Note that a brief LRP deflection was present on no-go trials, even though no response (or EMG activity) was present on these trials. Negative is plotted upward. (Adapted with permission from Miller and Hackley, 1992. © 1992 American Psychological Association.)

Trials"). From approximately 200–400 ms poststimulus, the ERPs were more negative over the left hemisphere when the shape of the stimulus was consistent with the right-hand response than with the left-hand response, and the complementary pattern was observed at the right-hemisphere site. After 400 ms, however, the ERP waveform was slightly more negative in both hemispheres for right-hand responses relative to left-hand responses.

Panel C shows the waveforms for the Go and No-Go trials after the LRP was isolated by means of equation 2.1. For both Go and No-Go trials, the LRP began to deviate from baseline at approximately 200 ms. On Go trials, the LRP continued until the end of the recording epoch. On No-Go trials, in contrast, the LRP returned to baseline at approximately 400 ms. Thus, the brain began to prepare the response indicated by the shape of the letter on both Go and No-Go trials, even though no response was executed on No-Go trials. This provides what I believe to be ironclad evidence that, at least under some conditions, response systems receive partial information about a stimulus before the stimulus has been fully identified.

Larger Issues There are many unjustified conclusions that you might be tempted to draw from the waveforms shown in figure 2.3. First, you might suppose that subjects typically began preparing the response at approximately 200 ms poststimulus, the time point at which the LRP began to deviate from baseline. However, as discussed previously in this chapter, the onset of a response in an averaged ERP waveform reflects the earliest onset times, not the average onset times (see rule 5). Thus, the 200-ms onset latency of the LRP in figure 2.3C reflects the fastest trials from the fastest subjects. Similarly, you might be tempted to assume that the LRP was only about half as large on No-Go trials as on Go trials. However, it is possible that the single-trial LRPs were just as large on No-Go trials as on Go trials, but were present on only 50 percent of trials.

Fortunately, the main conclusions from this experiment do not depend on any unjustifiable conclusions. In fact, it doesn't even matter whether or not the waveforms shown in figure 2.3C reflect the same LRP component that was observed in previous experiments. The simple fact that the hemispheric distribution of the voltage was different when the stimulus signaled a contralateral response rather than an ipsilateral response is sufficient to provide solid evidence that the brain had begun to determine which response was associated with the shape of the letter. So it doesn't matter if the effect reflects an ipsilaterally larger P3 component, a contralaterally larger N400 component, or some new, never-before-observed component (see strategy 6).

I would like to make one additional observation about this experiment: the data in figure 2.3 are extremely clean. In the raw ERP waveforms (panels A and B), the waveforms are almost completely noise-free during the prestimulus interval (the noise level in the prestimulus interval is very useful for assessing the quality of an ERP waveform). Even in the difference waves (panel C), which are plotted at a higher magnification, the prestimulus noise level is very small compared to the size of the LRP effects. Clean data lead to much greater confidence and much stronger conclusions; even if the p-value of an experimental effect passes the magical .05 criterion, noisy looking waveforms make a poor impression and make it difficult to have confidence in the details of the results (e.g., the onset time of an effect).

The following chapters will provide many hints for recording clean data, but I want to emphasize one factor here: clean ERP waveforms require a very large number of trials. For example, I usually try to have ten to twenty subjects in an experiment, and for each type of trial in the experiment I collect approximately sixty trials per subject when I'm examining a large component (e.g., P3 or N400), about 150 trials per subject for a medium-sized component (e.g., N2), and about 400 trials per subject for a small

Box 2.2 My Favorite ERP Component

Although I have never conducted an LRP experiment, I think the LRP may be the single most useful ERP component for addressing a broad spectrum of cognitive issues at this time. First, the LRP is very easy to isolate from other ERP components by means of difference waves. Second, it has been very well characterized, including its neural generator sources (for a review, see Coles, 1989). Third, it is relatively easy to design an experiment in such a way that you can isolate the LRP component (all you really need is to have some stimuli indicate left-hand responses and others indicate right-hand responses, with the assignment of stimuli to hands counterbalanced across trial blocks). Finally, the LRP has been used to provide solid answers to a variety of interesting questions (e.g., de Jong et al., 1990; Dehaene et al., 1998; Miller & Hackley, 1992; Osman & Moore, 1993) (see also review by Coles et al., 1995).

component (e.g., P1). This usually requires a session of three to four hours (including the time required to apply the electrodes and train the subject to perform the task). The need for large numbers of trials in ERP experiments is unfortunate, because it limits the questions that this technique can realistically answer. However, there isn't much point in conducting an experiment that would be very interesting except that the results are so noisy that they are difficult to interpret.

Example 3: Dual-Task Performance

Background It is often difficult for a person to perform two tasks at the same time. For example, it is difficult to discuss the principles of ERP experimental design while driving a BMW 325xi at high speeds on a winding road in a snowstorm. Dual-task interference can also be seen in much simpler tasks, and many researchers have recently studied dual-task interference in the *attentional blink* paradigm (see Shapiro, Arnell, & Raymond, 1997 for a review). In this paradigm, a very rapid stream of approximately twenty stimuli is presented at fixation on each trial, and the subject is required to detect two targets (called T1 and T2) from among the nontarget stimuli. Figure 2.4A illustrates a typical experiment. In this experiment, T1 and T2 are digits and the nontarget stimuli are letters. At the end of each trial, subjects report the identities of the two digits. While subjects are processing T1, they may be unable to effectively process T2, leading to errors in identifying T2. This would be expected to occur primarily when T2 is presented shortly after T1, and to assess the time course of the interference between T1 and T2, the lag between T1 and T2 is varied.

Figure 2.4B shows the pattern of results that is typically observed in this paradigm. T2 accuracy is typically severely impaired when T2 occurs two to four items after T1, but T2 accuracy is quite high at a lag of one item or at lags of five or more items. In contrast, T1 accuracy is typically quite good at all lags. This pattern of results is

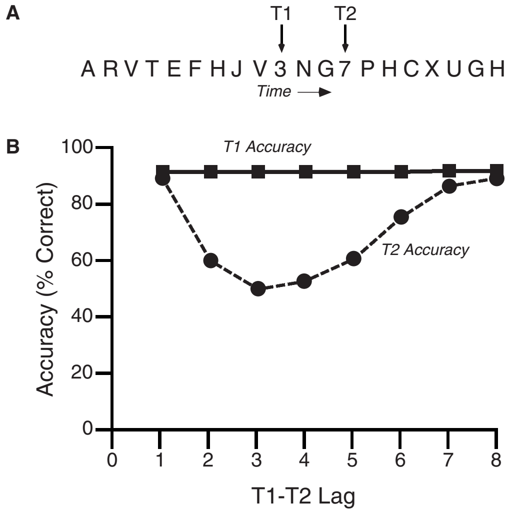

Figure 2.4 Experimental paradigm (A) and idealized results (B) from a typical attentional blink paradigm (based on the experiments of Chun & Potter, 1995). The stimuli are presented at fixation at a rate of ten per second. T1 and T2 are digits, and the other stimuli are letters; subjects are required to report the identities of T1 and T2 at the end of the trial. The lag between T1 and T2 is varied, and although T1 accuracy is generally found to be independent of lag, T2 accuracy typically drops significantly at lags 2–4.

called the *attentional blink* because it is analogous to the impairment in accuracy that would result from a T1-triggered eyeblink. This is an interesting example of dual-task interference, and it has generated a great deal of research in recent years. One of the most significant questions is whether the attentional blink reflects a failure to perceive T2 (Raymond, Shapiro, & Arnell, 1992) or whether T2 is perceived but is not stored in a durable form in working memory (Chun & Potter, 1995; Shapiro, Raymond, & Arnell, 1994). Potter (1976) demonstrated that stimuli can be identified more rapidly than they can be stored in working memory, and so it seemed plausible that both T1 and T2 could be identified even though only T1 could be stored in memory when T2 occurred shortly after T1.

To test this hypothesis, Ed Vogel, Kim Shapiro, and I conducted several experiments in which we examined the P1, N1, P3, and N400 components in variations on the attentional blink paradigm (Luck, Vogel, & Shapiro, 1996; Vogel, Luck, & Shapiro, 1998). Here, I will discuss only the most definitive experiment, in which we focused on the N400 component. The N400 component is typically elicited by words that mismatch a previously established semantic context. For example, a large N400 would be elicited by the last word of the sentence, ''I opened the dishwasher and pulled out a clean eyebrow,'' but a small N400 would be elicited if this same sentence ended with the word ''plate.'' The N400 can also be elicited with simple word pairs, such that the second word in PIG-HAT will elicit a large N400 whereas the second word in COAT-HAT will elicit a small N400. The N400 component is well suited for determining whether a word has been identified, because a semantically mismatching word cannot elicit a larger response unless that word has been identified to the point of semantic (or at least lexical) access. Thus, if a given word elicits a larger N400 when it mismatches the semantic context than when it matches the semantic context, this can be taken as strong evidence that the word was identified to a fairly high level. We therefore designed an experiment to determine whether the N400 component would be suppressed for words presented during the attentional blink period. Many previous experiments have examined the effects of semantic mismatch, so we were confident that we could adapt this approach to the attentional blink context (see strategy 2).

Experimental Design Figure 2.5A illustrates the stimuli and task for this experiment. Each trial began with the presentation of a 1,000-ms ''context word'' that established a semantic context for that trial. After a 1,000-ms delay, a rapid stream of stimuli was presented at fixation. Each stimulus was seven characters wide. Distractor stimuli were seven-letter sequences of randomly-selected consonants. T1 was a digit that was repeated seven times to form a seven-character stimulus. T2 was a word, presented in red, that

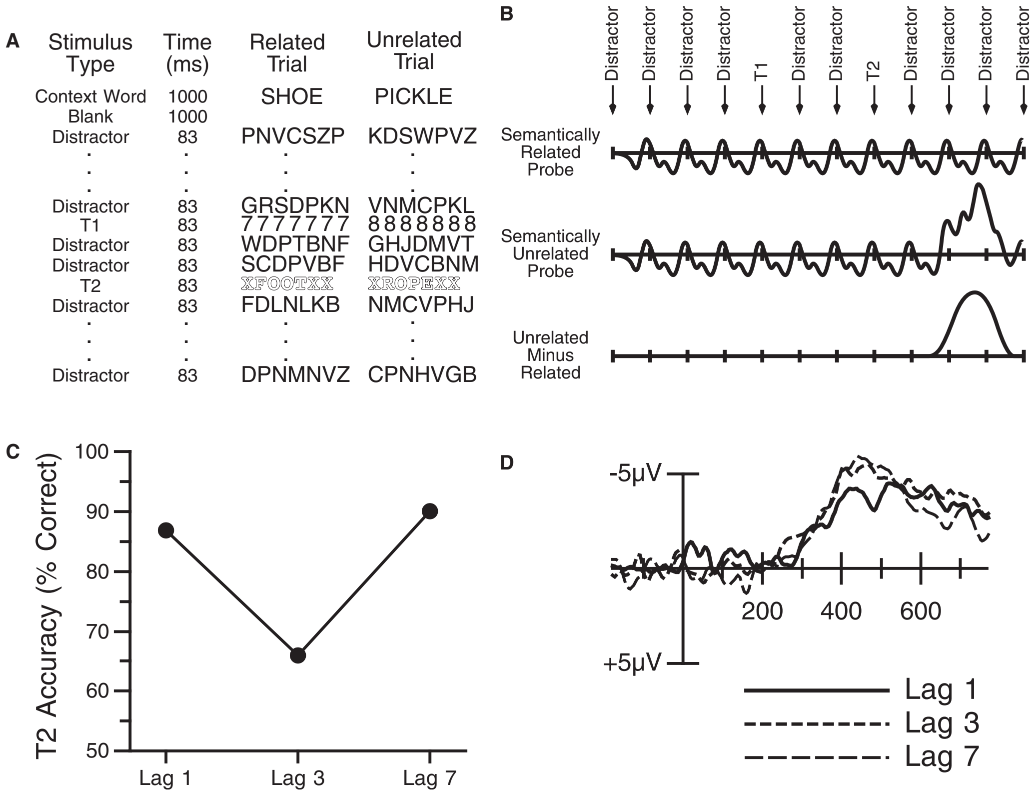

Figure 2.5 Paradigm and results from the study of Vogel, Luck, & Shapiro (1998). (A) Example stimuli. (B) Subtraction method used to overcome the overlap problem. (C) Mean discrimination accuracy for T2 as a function of lag. (D) Grand average ERP difference waveforms from the Cz electrode site, formed by subtracting related-T2 trials from unrelated-T2 trials. Negative is plotted upward.

was either semantically related or semantically unrelated to the context word that had been presented at the beginning of the trial. The T2 word was flanked by Xs, if necessary, to ensure that it was seven characters long. At the end of each trial, the subjects made two responses, one to indicate whether T1 was an odd digit or an even digit, and another to indicate whether T2 was semantically related or unrelated to the context word. Related and unrelated T2 words occurred equally often (as did odd and even T1 digits), and related words were very highly related to the context word (e.g., CAT-DOG). Each T2 word was presented twice for each subject (in widely separated trial blocks), appearing once as a related word and once as an unrelated word. This made it possible to be certain that any differences in the ERPs elicited by related and unrelated words were due to their semantic relationship with the context word rather than any peculiarity of the words themselves (see rule 6 and the Hillyard Principle). The lag between T1 and T2 was either 1, 3, or 7; this restricted set of lags was necessary so that a sufficient number of trials could be obtained at each lag.

As figure 2.5B illustrates the rapid presentation of stimuli in the attentional blink paradigm leads to an overlap problem when recording ERPs. Specifically, the response to a given stimulus lasts for hundreds of milliseconds, overlapping the responses to several of the subsequent stimuli. This makes it difficult to isolate the ERP elicited by T2 from the ERPs elicited by the preceding and subsequent stimuli. To overcome this problem, we computed difference waves (see strategy 4) in which we subtracted the response elicited by T2 when it was semantically related to the context word from the response elicited by T2 when it was an unrelated word (see figure 2.5B). The responses to the other items in the stimulus stream should be essentially identical on related-T2 and unrelated-T2 trials, and this difference wave therefore provides a relatively pure measure of the brain's differential response as a function of the semantic relatedness of T2 relative to the context word. A large difference between related-T2 and unrelated-T2 trials can therefore be used as evidence that the T2 word was identified sufficiently to determine its semantic relationship to the context word.

Results Panels C and D of figure 2.5 show the results of this experiment. Accuracy for reporting whether T2 was semantically related or unrelated to the context word was highly impaired at lag 3 relative to lags 1 and 7; this is the usual attentional blink pattern. In contrast, the N400 component elicited by T2 was equally large at all three lags. Thus, although the subjects could not accurately report whether T2 was related or unrelated to the context word at lag 3, the N400 wave differentiated between related and unrelated trials, indicating that the brain made this discrimination quite accurately. This result indicates that stimuli are fully identified during the attentional blink, but are not reported accurately because they are not stored in a durable form in working memory.

This conclusion relies on a hidden assumption, namely that the N400 component would be significantly smaller if the perceptual processing of T2 had been impaired at lag 3. This is a problematic assumption, because it is difficult to know what the relationship is between the amplitude of an ERP component and the speed or accuracy of a cognitive process. This brings us to another rule:

Rule 11. Never assume that the amplitude and latency of an ERP component are linearly or even monotonically related to the quality and timing of a cognitive process. This can be tested, but it should not be assumed.

To avoid violating this rule, we conducted a control experiment to determine whether a decrease in the perceptibility of a word would cause a significant decrease in N400 amplitude. In this experiment, we simply added random visual noise of varying intensity to the words. As the intensity of the noise increased, both behavioral accuracy and N400 amplitude declined in a roughly linear manner. Moreover, a change in accuracy that was comparable to the impaired accuracy observed at lag 3 in the main experiment led to a large and statistically significant decline in N400 amplitude. Thus, the absence of a decline in N400 amplitude at lag 3 in the main experiment provides strong evidence that the behavioral errors at lag 3 reflected an impairment that followed word identification.

These results also suggest that a great deal of processing can occur in the absence of awareness, because the brain apparently identified the T2 word at lag 3 even though subjects could not report the word (although it is possible that subjects were briefly aware of the T2 word even though they could not report its semantic relationship moments later).

Larger Issues I would like to draw attention to three aspects of this attentional blink experiment. First, this experiment used ERPs to monitor a psychological process—word identification—that we suspected was present but could not be observed in the subjects' overt behavior. Many ERP experiments are designed to show correspondences between behavioral results and ERPs, but it is often more interesting to demonstrate an interesting pattern of both similarities and differences. I should note that we ran an additional attentional blink experiment that focused on the P3 wave, and we found that the P3 wave was completely suppressed at lag 3, consistent with the proposal that the attentional blink reflects an impairment in working memory. In this manner, we were able to show a theoretically sensible pattern of similarities and differences between overt behavior and ERPs.

A second notable aspect of this experiment is that we used the N400 component as an index of word identification even though the neural/psychological process that generates the N400 component is probably not directly related to word identification. Because word identification was a necessary antecedent to the presence of an N400 in this experiment, however, we could use the N400 as an indirect index of word identification. Similarly, the P3 wave can be used as an index of ‘‘stimulus evaluation time’’ even though the neural/psychological process that generates the P3 component is probably unrelated to perception. For example, because improbable stimuli elicit a larger P3 than probable stimuli, it is possible to isolate the P3 wave by constructing improbable-minus-probable difference waves. Because the brain cannot produce a larger P3 for an improbable stimulus until that stimulus has been identified and categorized, the timing of the P3 wave in an improbable-minus-probable difference wave can be used as an indirect index of identification and categorization (see Donchin, 1981; Kutas, McCarthy, & Donchin, 1977; McCarthy & Donchin, 1981; Vogel & Luck, 2002).

If you look at ERP experiments that have had a broad impact in cognitive psychology or cognitive neuroscience, you will find that many of them use a given ERP component that is not obviously related to the topic of the experiment. For example, our attentional blink experiment used the language-related N400 component to examine the role of attention in perceptual versus post-perceptual processing. Similarly, Dehaene et al. (1998) used the motor-related LRP component to address the possibility of perception without awareness. One of my graduate students, Adam Niese, refers to this as *hijacking* an ERP component, and it leads to an additional strategy of ERP experimental design.

Strategy 7. Hijack Useful Components from Other Domains

It is useful to pay attention to ERP findings from other subareas of cognitive neuroscience. A component that arises ‘‘downstream’’ from the process of interest may be used to reveal the occurrence of the process of interest.

A third notable aspect of this experiment is that difference waves were used to isolate both the activity elicited by a single stimulus and a specific ERP component elicited by that stimulus (see strategy 4). It is often necessary to present stimuli in such close temporal proximity that the ERPs elicited by each stimulus will overlap each other in time, and difference waves can often be used to circumvent this problem (for additional examples of this approach, see Luck, 1998b; Luck, Fan, & Hillyard, 1993; Luck & Hillyard, 1995). You must use care, however, because you cannot simply assume that the difference wave provides a pure estimate of the size of the component of interest. In the attentional blink experiment, for example, the overall N400 may have been smaller at lag 3 than at lags 1 and 7, even though the difference in N400 amplitude between related and unrelated T2 words was the same at all three lags. In this experiment, the key question was whether the brain could differentiate between related and unrelated T2 words, and the size of the overall N400 was immaterial. In fact, it didn't even matter whether the activity in the unrelated-minus-related difference waves was the N400, the P3, or some other component; no matter what component it was, it demonstrated that the brain could distinguish between related and unrelated words (see strategy 6). Thus, difference waves can be very useful, but they must be interpreted carefully.

Suggestions for Further Reading

There are many examples of clever and thoughtful ERP experimental designs in the literature. The following is a list of a few of my favorites. I would definitely recommend that new ERP researchers study these experiments carefully.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*, 597–600.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error-detection and compensation. *Psychological Science*, *4*, 385–390.

Gratton, G., Coles, M. G. H., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and post-stimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 331–344.

Handy, T. C., Solotani, M., & Mangun, G. R. (2001). Perceptual load and visuocortical processing: Event-related potentials reveal sensory-level selection. *Psychological Science*, *12*, 213–218.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, *21*, 171–186.

Paller, K. A. (1990). Recall and stem-completion priming have different electrophysiological correlates and are modified differentially by directed forgetting. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 1021–1032.

Van Petten, C., & Kutas, M. (1987). Ambiguous words in context: An event-related potential analysis of the time course of meaning activation. *Journal of Memory & Language*, *26*, 188–208.

van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, *280*, 572–574.

Winkler, I., Kishnerenko, E., Horvath, J., Ceponiene, R., Fellman, V., Huotilainen, M., Naatanen, R., & Sussman, E. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences*, *100*, 11812–11815.

Woldorff, M., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, *79*, 170–191.

Summary of Rules, Principles, and Strategies

Rule 1. Peaks and components are not the same thing. There is nothing special about the point at which the voltage reaches a local maximum.

Rule 2. It is impossible to estimate the time course or peak latency of a latent ERP component by looking at a single ERP waveform—there may be no obvious relationship between the shape of a local part of the waveform and the underlying components.

Rule 3. It is dangerous to compare an experimental effect (i.e., the difference between two ERP waveforms) with the raw ERP waveforms.

Rule 4. Differences in peak amplitude do not necessarily correspond with differences in component size, and differences in peak latency do not necessarily correspond with changes in component timing.

Rule 5. Never assume that an averaged ERP waveform accurately represents the individual waveforms that were averaged together. In particular, the onset and offset times in the averaged waveform will represent the earliest onsets and latest offsets from the individual trials or individual subjects that contribute to the average.

Rule 6. Whenever possible, avoid physical stimulus confounds by using the same physical stimuli across different psychological conditions (the Hillyard Principle). This includes ''context'' confounds, such as differences in sequential order.

Rule 7. When physical stimulus confounds cannot be avoided, conduct control experiments to assess their plausibility. Never assume that a small physical stimulus difference cannot explain an ERP effect.

Rule 8. Be cautious when comparing averaged ERPs that are based on different numbers of trials.

Rule 9. Be cautious when the presence or timing of motor responses differs between conditions.

Rule 10. Whenever possible, experimental conditions should be varied within trial blocks rather than between trial blocks.

Rule 11. Never assume that the amplitude and latency of an ERP component are linearly or even monotonically related to the quality and timing of a cognitive process. This can be tested, but it should not be assumed.

The Hillyard Principle: Always compare ERPs elicited by the same physical stimuli, varying only the psychological conditions.

Strategy 1. Focus on a specific component.
Strategy 2. Use well-studied experimental manipulations.
Strategy 3. Focus on large components.
Strategy 4. Isolate components with difference waves.
Strategy 5. Focus on components that are easily isolated.
Strategy 6. Use component-independent experimental designs.
Strategy 7. Hijack useful components from other domains.