

## Chapter 15

### Reading, Writing, and Reviewing ERP Papers

#### Overview

This chapter will provide some advice about reading, writing, and reviewing ERP papers. I will try to limit myself to issues that are ERP-specific (or unusually common/important in ERP studies), although I will discuss a few broader issues near the end of the chapter. If you'd like to read a useful general discussion of strategies for writing journal articles, I highly recommend an old chapter by Daryl Bem (Bem, 1987). Bem became a controversial figure later in his career, but no matter what you think about his research, there is no doubt that he knew how to write an effective journal article.

The Society for Psychophysiological Research (SPR), which publishes the journal *Psychophysiology*, occasionally publishes publication guidelines for specific methods, including ERPs (Keil, Debener, Gratton, Junhöfer, Kappenman, Luck, Luu, Miller, & Yee, 2014; Picton, Bentin, Berg, Donchin, Hillyard, Johnson, Miller, Ritter, Ruchkin, Rugg, & Taylor, 2000). These are very useful to read, and several of the specific recommendations that I will make in this chapter can also be found in these guidelines. This shouldn't be a surprise, because Emily Kappenman and I were on the committee that wrote the most recent set of guidelines. This version also includes a checklist of methodological details that should be found in every ERP paper (unless the authors make a good argument for excluding a given piece of information). All papers published in *Psychophysiology* are supposed to follow the publication guidelines and checklist when submitting a manuscript. I would strongly recommend that you make a copy of the checklist and use it every time you are getting ready to submit a manuscript and any time you are reviewing an ERP manuscript, no matter what the journal is.

#### The Journal

A journal article must match the goals and audience of the journal in which it's published. This influences how you will write your papers, and it is an important criterion when you are reviewing a paper submitted to a specific journal. As a reviewer and editor, I have often encountered submissions that were simply inappropriate, either in content or in style, for the journal, but would be completely appropriate for a different journal. For example, papers published in most content-oriented journals (e.g., *Cognitive Psychology*, *Journal of Neuroscience*, *Archives of General Psychiatry*) must make it clear how they advance the field in that content area and not focus on *ERPology* (the study of ERPs). They must also be written in a way that readers who are not very ERP-savvy can easily understand the essence of the methods and results. In contrast, papers written for method-oriented journals (e.g., *Psychophysiology*, *NeuroImage*) can focus on ERPology and should assume an ERP-savvy readership.

#### The Introduction

The Introduction, more than any other part of a manuscript, must be carefully tailored for the intended journal and audience. If you do a good job with the Introduction, the rest of the paper will be easy for the reader to understand and evaluate. If you do a poor job with the Introduction, then the quality of your data and analyses will not be appreciated, and your paper will not live up to its potential.

When I teach undergraduate students how to read a journal article, I start by making a distinction between the *Big Hypothesis* that is tested by the article and the *predictions* of the specific experiment. These two things are usually conflated in courses on research methods, which focus on a null hypothesis and an alternative hypothesis. However, the Big Hypothesis is very different from the null and alternative hypotheses in most real journal articles. The Big Hypothesis is something that could be tested with many different kinds of experiments. For example, a Big Hypothesis might be something like "the response selection process is slowed in people with schizophrenia." This hypothesis could be tested in many ways. One experiment testing this Big Hypothesis using the P3 wave and the lateralized readiness potential (LRP) was briefly described in Chapter 1. Within the context of this specific experiment, we predicted that P3 latency would not be slowed in patients but that LRP onset latency would be slowed. Thus, the Big Hypothesis was about response selection processes (a

relatively abstract and theoretical entity), whereas the specific predictions were about P3 and LRP latencies (concrete variables that can be measured and subjected to statistical analyses).

Most good ERP papers describe both a Big Hypothesis and specific predictions. Indeed, the main goal of the Introduction section of an ERP paper is to provide the reader with the logical “glue” that links the Big Hypothesis to the general design of the experiment and then to the specific predictions. This typically involves three steps:

1. Describe the broad theoretical background for the study, motivating the reader to care about the Big Hypothesis. If you can't make the readers care about the Big Hypothesis, the paper is pointless.
2. Describe the general methodological approach of the present study. Specifically, what are the key elements of your methods that will allow your study to discover something new? In an ERP paper, this often includes a discussion of why ERPs will be particularly useful in addressing the Big Hypothesis, and it might also describe the specific ERP components or effects that are central to your methodological approach.
3. Describe the specific predictions of the study. In other words, if the Big Hypothesis is correct, then what pattern of amplitude and/or latency effects should be observed for the specific ERP components that you will be measuring? You should typically be very specific. Don't just say that an experimental manipulation will affect Component X; instead, say that the amplitude of Component X will be greater in Condition A than in Condition B (or that the latency will vary, or whatever it is that you're predicting). It's also a good idea to indicate some alternative patterns of results that might be obtained and how they would disconfirm your Big Hypothesis. It's even better if you can contrast the predictions that arise from your Big Hypothesis with the predictions of some other plausible hypothesis. If your paper features a battle between competing hypotheses, it will be more interesting and have a bigger impact.

When you read the Introduction section of an ERP paper, I encourage you to focus on the logic that leads from the Big Hypothesis to the general methodological approach and then to the specific predictions. If you are reviewing a journal submission, you should insist that the authors provide this logic and that they provide both a Big Hypothesis and specific predictions (see the list of Top Ten Reasons to Reject an ERP Manuscript at the end of this chapter). Providing specific predictions about the ERP measures is actually the first item on the *Psychophysiology* checklist.

I was an associate editor at *Cognitive, Affective, and Behavioral Neuroscience* (CABN) for many years, and I handled most of the ERP studies. After about a year, I realized that a lot of reasonably good studies required two or three rounds of revisions before being accepted as a result of an ineffective Introduction. Because the authors had not adequately described the logic of the study and had not provided detailed predictions, the reviewers did not really understand how the conclusions followed from the results and why the conclusions were important (even when the conclusions were, in fact, valid and important). As a result, I began checking the Introduction of each manuscript for specific predictions before sending it out for review. If the manuscript did not lay out the logic of the study and provide specific predictions, I would send it back to the authors for revisions prior to sending it out for review. This saved time (and grief) for the authors, because they avoided the round of negative reviews that inevitably occurred for a manuscript that lacked these elements. This saved time for the reviewers, because they didn't have to read a hard-to-understand manuscript before getting a revision that had an appropriate Introduction. And it saved time for me, because I didn't have to wade through a bunch of negative reviews trying to determine if the manuscript could be salvaged by a revision. Box 15.1 provides an example of the kind of decision letter I would write when I received a manuscript that lacked specific predictions. As you will see, it also contains some advice about the Discussion section and about framing the study so that it will be of interest to a broad audience.

*Box 15.1- Example Editorial Decision Letter*

Dear Dr. XXXXX:

I gave your manuscript a quick reading so that I could choose appropriate reviewers. After this quick reading, it was clear to me that the manuscript cannot be accepted in its present form. Thus, to save everyone time and effort, I am rejecting this version of the manuscript, but I will allow you to submit a revised manuscript if you can overcome the problems with the current version.

In most ERP papers that make a significant contribution to broad questions in cognitive neuroscience, the Introduction ends with a set of specific predictions about the pattern of ERP results that will be obtained. That is, competing hypotheses about broad issues in cognitive neuroscience are raised in the first part of the Introduction, and then specific predictions about the ERP results are given that will distinguish between the competing hypotheses. This is both an indicator of the likely importance of the study and an important aid to the readers, who can much better understand the methods and results if they know what general patterns of results can be expected and how these results are related to the broad issues addressed by the study. The present manuscript does not contain such predictions. If you submit a revision, you should rewrite the Introduction in this manner.

It is also useful for the Discussion section to begin with a brief recap of the major findings and the conclusions that can be drawn from them. Again, this helps make it clear that the study represents a significant advance in knowledge about the broad issues that were initially raised in the Introduction. It is also helpful for the casual reader who does not want to read the details of your methods and results, but instead wishes to quickly see your “bottom line.” It was quite difficult for me to determine what conclusions could be drawn from this manuscript. If you submit a revision, you should rewrite the beginning portion of the Discussion in this manner.

## **The Methods**

In this section, I'll start by describing some of the most common methodological issues that I encounter when reading and reviewing ERP papers (many of which are in the list of Top Ten Reasons to Reject an ERP Manuscript at the end of the chapter). When you are reading/reviewing an ERP paper, you should specifically look for these issues. And when you are writing an ERP paper, you should make sure that these issues are adequately addressed in your own methods.

A general issue about ERP processing and analysis procedures is that every procedure has some chance of being inappropriate or creating an error, and the more steps of processing/analysis are applied, the greater is the chance that the results will be wrong or will lead to incorrect conclusions. Some processing steps (e.g., source analysis) are more likely to lead you astray than others (e.g., the application of mild filters). Moreover, many of these steps give the researcher an opportunity to set parameters that have a large impact on the results and might intentionally or unintentionally bias the results to fit the researcher's expectations. As a result, you should be skeptical of the conclusions of a paper that involves a long and complicated set of processes, especially when those processes give the researcher a lot of control over how the results turn out. If the data are tortured sufficiently, they may end up saying whatever the researcher wants to hear!

### *Filter Settings and Reference*

The first things I look for in a Methods section are the filter settings and the location of the reference electrode. Until you know how the data have been filtered and referenced, you can't know how to interpret the ERP waveforms. If the study uses the average of all sites as the reference, you should check to see whether the electrodes are broadly and evenly distributed over the scalp. If not, then the average reference may be inappropriate. And if the manuscript uses the average reference and doesn't provide a list of all the electrode sites, then that's a problem. As discussed in detail in Chapter 5, the reference contributes just as much to the

ERP waveform as the active site, so it's hard to evaluate an ERP waveform if you don't know which sites were used to create the reference portion of the signal. In general, any reference is potentially legitimate, but it can be a problem if the reference is different from what is usually used in a given area of research because it will be hard to compare waveforms across studies. Occasionally, a paper will report an unusual scalp distribution or even claim to have discovered a new component because of the use of a different reference. As an editor, I have occasionally asked authors to provide a set of figures in online supplementary materials that show the waveforms referenced in a more typical manner. You should feel free to request this as a reviewer.

When you are looking at the filter settings in an ERP paper, the first thing you want to determine is whether the filtering is so severe that it might be causing artifactual effects or substantial changes in onset or offset latencies (see, e.g., Figure 7.10 in Chapter 10 and Figures 12.8-12.11 in Online Chapter 12). As a reviewer, I will not typically raise questions about filtering unless the high-pass cutoff is  $\geq 0.5$  Hz or the low-pass cutoff is  $\leq 10$  Hz. But if the bandpass is any narrower than this, it is quite possible that the waveforms were distorted by the filters. If you are reviewing a manuscript with a narrow bandpass, it is perfectly reasonable for you to ask the authors to provide a figure showing the effects of the filter on artificial waveforms. Indeed, one of my grad students recently received a review asking for such a figure (even though our half-amplitude cutoffs were 0.1 and 30 Hz, with a gentle slope of 12 dB/octave, which should produce very little distortion in most cases).

### *Stimuli and Task*

You should take a close look at all the stimulus and timing parameters. At a minimum, you should be able to reconstruct the stimuli and timing from the description in the Methods section. This is one of those things that seems obvious but is still a common problem. I also recommend including a figure that shows example stimuli and the timing, even for very simple experiments. This will make it easier for someone to skim your paper and understand your task.

If the study is looking at relatively early ERP components ( $\leq 200$  ms), the results could easily be distorted by confounds in the stimuli. In many cases, studies simply fail to counterbalance the stimuli (e.g., using one set of stimuli in one condition and a different set of stimuli in another condition), which is an obvious confound. In other cases, the stimuli have been counterbalanced but confounds arise from the order of stimuli (e.g., differential adaptation because of differences in stimulus repetition between conditions). Chapter 4 describes both the obvious and subtle confounds in detail, but they can be summed up with the Hillyard Principle: To avoid sensory confounds, you must compare ERPs elicited by exactly the same physical stimuli, varying only the psychological conditions. The best way to make sure that the Hillyard Principle is being followed is to ask whether the experiment uses exactly the same stimulus sequences in every condition, varying only the instructions<sup>1</sup>. If not, then there might be a subtle stimulus confound.

If a study doesn't counterbalance the stimuli or suffers from a subtle sensory adaptation confound, this may be a minor problem or it may be a major problem. It's typically a major problem if the study hinges on comparisons of sensory responses across conditions (e.g., activity within the first ~150 of stimulus onset). However, sensory confounds are typically a minor problem if the study focuses on a late component such as P3 or N400. In the latter case, it's usually sufficient for the paper to make mention of the potential confound, either in a footnote or in the Discussion section. But in the former case, this might be grounds to reject the manuscript (or to disbelieve the conclusions if the paper has already been published).

It's also important to look at the details of the stimulus timing to see if overlapping activity from the previous trial is likely (see Online Chapter 11 for a detailed discussion). Overlap-related problems primarily

---

<sup>1</sup> To be more precise, you want to make sure that the stimulus sequences are generated following exactly the same *rules* in each condition, with only the instructions varying systematically across conditions. You won't usually want to use precisely the same stimulus sequences in multiple conditions, because this might give subjects an opportunity to learn the sequences. But you do want to make sure that any differences in the sequences used across conditions are purely random rather than differing systematically.

arise when the overlap differs across conditions, producing a confound. Note, however, that even nondifferential overlap can distort scalp distributions and source localization (see Figure 11.4 in Chapter 11).

A common overlap-related issue occurs when a fixation point appears at the beginning of each trial (e.g., 300 ms before a target stimulus), creating an ERP that overlaps with the baseline period of the subsequent stimuli. This causes the baseline to look funny (when time-locked to the target). Both authors and readers of journal articles often fail to think of the fixation point as a “real stimulus,” so they don’t realize that it generates a substantial ERP that distorts the data from the subsequent stimulus. Overlap from a fixation point is usually the same across conditions and doesn’t create a confound, but the manuscript needs to explain why the baseline looks funny in this situation. Similarly, the offset of a stimulus immediately prior to the beginning of a trial can create an ERP that contaminates the baseline.

An ERP paper should also provide clear information about the number of trials that contributed to each ERP waveform. In many cases, this information is provided only indirectly (e.g., “there were X total trials in each of Y blocks, with 25% being stimulus A and 75% being stimulus B”). The reader should not have to do math to figure out the number of trials that were averaged together; this information should be made explicit (e.g., “after collapsing across the counterbalancing factors, there were 200 trials in the averaged ERP waveform for condition A and 600 trials for condition B”).

If the waveforms being compared are based on different numbers of trials, this is not a problem in some cases but is a significant confound in others. This issue is described in detail in the online supplement to Chapter 9 (available at <http://mitpress.mit.edu/luck2e>). Briefly, if a linear measure such as mean amplitude is used to quantify the ERP amplitudes, then it’s not usually a problem to have different numbers of trials. If a nonlinear measure such as peak amplitude is used, however, the measured values will be biased to be larger in the waveforms with fewer trials. This is typically OK if the study finds a larger amplitude in the condition with more trials (because that’s the opposite of the effect that would be produced by the confound), but it can be a significant problem if the condition with fewer trials is also the condition with a larger amplitude.

### *State-Related Confounds*

A common confound in ERP studies is that subjects are in different states of arousal or preparation across groups or conditions. This can usually be prevented by using within-subject designs and randomly intermixing the different trial types, and this approach should be used whenever possible. However, some kinds of experimental manipulations cannot be randomly intermixed, and studies that compare different types of subjects are by definition between-subject designs. When you are reading or reviewing a paper that uses a between-subjects design or does not randomly intermix the trial types in a within-subjects design, you should consider whether any differences in the ERPs could be a direct or indirect consequence of differences in state.

A common version of this problem arises when one condition of an experiment is more difficult than another, and the level of arousal might therefore be greater in the more difficult condition. *Arousal* is not a well-defined, unitary construct, and it is difficult to know how arousal might impact a given ERP component. However, increased arousal typically leads to increases in sensory responsiveness. For example, the P1 wave becomes larger when subjects are instructed to respond very rapidly (see, e.g., Figure 3.2 in Chapter 3). Cognitive factors do not typically influence the earliest visual components (C1 and P1) or the earliest auditory components (the ABRs and midlatency components), so if you see an effect of a cognitive manipulation on these components, you should ask whether they could be explained by differences in arousal. However, later components might also be influenced by differences in arousal or effort. For example, P3 amplitude is modulated by the amount of resources allocated to the task (see Chapter 3).

Differences in state or preparation can also lead to baseline differences, which are then propagated to the poststimulus waveform by the baseline correction process (see Figures 8.1 and 8.2 in Chapter 2). For example, if the subject has greater ability to anticipate the time or identity of a stimulus in one condition than in another, this may lead to differences in the CNV during the prestimulus interval, which may then distort the entire waveform. Effects of this nature often lead to very slow effects that, because of baseline correction, appear

very early in the ERP waveform (within the first 100 ms) and persist for hundreds of milliseconds (as in Figure 8.1).

### *Artifact Rejection and Correction*

It has become common for ERP papers to provide very little information about how artifacts were treated (e.g., “trials with artifacts were excluded from the averaged ERP waveforms” or “ICA was used to perform artifact correction”). An ERP paper should provide enough information that you could replicate the artifact rejection/correction procedures. For artifact rejection, this includes describing the algorithm that was used to define an artifact. If artifacts were rejected manually, the manuscript should indicate whether the person doing the rejection was blind to group membership (for between-group designs). For artifact correction, this includes describing how many components were removed, the procedure for determining which components should be removed, and whether the person making decisions about component removal was blind to group membership.

In addition, when artifact rejection is used, the paper should describe the average percentage of trials that were rejected and the range across subjects (e.g., “On average, 18.4% of trials were excluded from the averages because of artifacts (range = 12.1-23.4%)”). If some subjects are excluded from analysis because of excessive artifacts, the paper should say what the criterion was for excluding subjects and whether this criterion was a priori or post hoc. Papers from my lab typically say something like this: “We always exclude subjects for whom more than 25% of trials were rejected; 2 subjects were excluded for this reason in the present experiment.”

Exclusion of subjects makes me skeptical if it isn't done on the basis of an a priori criterion. I particularly hate to read something like “5 of the 18 subjects were excluded because of excessive artifacts or poor data quality” (with no more information about exactly why these subjects were excluded and the remaining subjects were included). This makes me think that the effects were not significant in the initial statistical analyses, and the researchers started excluding subjects in hopes of finding significant effects. This approach leads to a high likelihood of bogus effects (Type I errors). If you read a published paper that excludes more than 1-2 subjects without a detailed description of why they were excluded, you should be skeptical of the results. And if you are a reviewer of a journal submission, you should make the authors provide a detailed justification for excluding subjects.

Failure to adequately exclude trials with blinks and/or eye movements is a common source of confounds in ERP studies. As discussed in Chapter 6, there are two types of problems that arise from ocular artifacts. First, there are the direct effects of the artifactual voltages on the ERP waveforms. Blinks are huge, so if 2% of trials are contaminated with stimulus-triggered blinks in one condition, and none have blinks in another condition, this could lead to a difference between conditions that is larger than many ERP effects, even at posterior electrode sites. Eye movements generate potentials that vary in amplitude with the size of the eye rotation and can also be very large.

Eye movements are especially problematic in visual experiments that involve lateral stimuli, which tend to trigger eye movements. Even with the most sensitive methods, it is difficult to reject eye movements that are smaller than about 1° of rotation, which correspond to an EOG voltage of approximately 16  $\mu\text{V}$ . Thus, even if a paper reports performing rigorous rejection of trials with eye movements, small eye movements may remain that are still large enough to produce a relatively large contamination of the ERPs. This is especially problematic in experiments that examine lateralized ERP components, such as N2pc, CDA, and LRP. Box 6.3 in Chapter 6 describes a method for ensuring that these components are not contaminated by significant EOG signals. If you read, write, or review papers that focus on these components, I highly recommend that you read about that method, use it yourself, and insist that other researchers use it as well.

The second problem with blinks and eye movements is that they change the sensory input. Consequently, if the eyes systematically differ in position across conditions, or if subjects are more likely to blink during the stimulus in some groups or conditions, the ERPs will differ across conditions because of a difference in sensory input. This is especially problematic in experiments with lateral stimuli that are designed in a manner that allows the subject to predict the location of a given stimulus, so that the eyes may move prior to stimulus onset.

EOG recordings are reasonably sensitive at detecting changes in eye position (of approximately 1° or more), but they do not typically provide information about absolute eye position. Thus, if the eyes move prior to the beginning of the epoch, the EOG signal may not indicate that an artifact is present during the epoch, and the trial will not be rejected. Preventing and assessing this possibility is quite difficult (unless a video-based eye tracker is used to record eye position concurrently with the EEG recording). The best solution is to make the stimulus locations unpredictable, so that gaze will be centered at the beginning of each trial.

As noted in Chapter 6, artifact correction can eliminate the direct electrical potentials produced by blinks and, to a more limited extent, eye movements. However, artifact correction cannot eliminate the effects of blinks or eye movements on the sensory input. Thus, it is often necessary to combine correction with rejection of artifacts that might change the sensory input. However, very few published papers actually do this. As a reviewer of a manuscript that uses artifact correction, you should demand that the manuscript rules out sensory changes owing to ocular artifacts. This can be done by rejecting trials with eye movements or blinks that occur within  $\pm 200$  ms of stimulus onset. However, some analysis software makes it difficult to combine rejection and correction in this manner. If blinks are the only likely source of ocular artifacts that might change the stimuli (e.g., because all the stimuli are presented at fixation), then it can be sufficient for the paper to demonstrate that very few blinks occurred near the time of stimulus onset.

### *Amplitude and Latency Quantification*

One of the greatest methodological challenges in most ERP studies is to figure out an optimal way to quantify the amplitudes and latencies of the ERP components in a way that both maximizes statistical power and avoids inflating the Type I error rate. These issues mainly arise in the choice of the time windows and electrodes sites. As discussed in detail in Chapter 10, the Type I error rate (i.e., the probability that bogus effects will be significant) may be dramatically increased if the observed waveforms are used to select the time windows and electrode sites for analysis. However, if the observed waveforms are not used to select the time windows and electrodes, power may be reduced because components may not be measured from the windows and electrodes where the effects are largest. Chapter 10 discusses several strategies that can be used to select time windows and electrodes in an unbiased manner while still maintaining reasonable power, and it is essential that all ERP studies use an unbiased approach (e.g., a priori selection based on previous studies).

Consequently, when you are reading/reviewing an ERP paper, you should look carefully at how the measurement windows are defined. In most cases, the paper should explicitly state the rationale for choosing the windows and sites (although sometimes this will be implicit). As a reviewer, you should demand that the authors provide a convincingly unbiased rationale. As a reader, you should be suspicious if no rationale is given and (a) the windows/electrodes seem very arbitrary (e.g., a window of 183-247 ms, or noncontiguous electrode sites), or (b) the windows/electrodes vary across experiments with no explanation.

You should also think carefully about the validity of the conclusions if the study quantifies amplitudes or latencies using peak measures. As described in Chapter 9, these measures can lead to a range of problems, ranging from low power to biased results. Peak measures aren't always a problem, but they can be in some cases, and you should think about whether they are appropriate for a given study. Of course, other measures can also be problematic, such as the strong dependence of mean amplitude measurements on the choice of the measurement window. But peak measures are much more likely to be problematic.

### *Statistical Analysis*

ERP studies lead to very rich data sets with millions of values that contribute to the results. Consequently, it is possible to conduct a large number of statistical analyses in any given experiment. The more analyses are conducted, the greater is the probability that one or more of the statistically significant results are bogus. Thus, try to limit the number of p values that you compute in your own studies, and be skeptical of the results of studies that involve computing a large number of p values. This can be true even when a correction is applied for multiple comparisons, because researchers usually correct on the basis of the number of comparisons for a single analysis, not the number of comparisons in the entire study. Moreover, when factorial ANOVAs are used, the number of p values increases exponentially as the number of factors increases (because of all the

possible interactions among factors), and almost no one corrects for this. Consequently, it is better to do statistical analyses after collapsing across as many factors as possible (and measuring from difference waves, which also reduces the number of factors in the ANOVA).

It can be difficult to assess this problem in the context of a submitted or published paper, because you have no way of knowing how many analyses were conducted but not reported. However, you should be skeptical when a paper reports a lot of  $p$  values (or clearly involved a lot of ANOVAs with multiple factors). As a reviewer, you could certainly say that a result should be replicated before being published. You can also ask the authors to re-do the analyses after collapsing across factors and after measuring the amplitudes or latencies from difference waves.

There are two technical factors that you should always look for in the statistical analysis of an ERP study (as described in Chapter 10). First, if any factors have more than two levels, a correction for heterogeneity should be applied (e.g., the Greenhouse-Geisser correction). Second, care must be taken in interpreting condition  $\times$  electrode interactions (and normalization procedures should be carefully scrutinized because of the problems raised by Urbach & Kutas, 2002, 2006).

## The Results and Figures

If you've carefully checked everything in the Methods section, the Results section should be fairly straightforward. Note, however, that authors often describe some of the methodological details in the Results section rather than in the Methods section. Although this may seem wrong, it's often a good idea because it means the reader will have those details fresh in the mind when reading the relevant part of the Results section.

### *Presentation of Statistical Analyses*

I have two specific recommendations about the presentation of statistical analyses. First, it can be really hard to read a Results section that has tons of  $F$ ,  $t$ ,  $r$ , and  $p$  values embedded in the text, but this is common in ERP studies that involve analyses of amplitudes and latencies of multiple components. The Results section can be made more readable by putting these values into one or more tables. This can really make life easier for the readers and reviewers of the paper. And you definitely want to make life easy for the reviewers, who hold your fate in their hands!

Second, the inferential statistics should ordinarily follow rather than precede the description of the actual data. Inferential statistics such as ANOVAs are used to tell us how much confidence we can have in our data. As such, they are not data but are instead a means of determining whether the data pattern is believable. However, many ERP results sections are written as if the inferential statistics are the primary results, with virtually no mention of the descriptive statistics or the ERP waveforms. For example, consider this excerpt from a Results section:

*There was a significant main effect of probability on P3 amplitude,  $F(1,9) = 4.57, p < .02$ . There was also a significant main effect of electrode site,  $F(11,99) = 3.84, p < .01$ , and a significant interaction between probability and electrode site,  $F(11,99) = 2.94, p < .05$ ...*

In this example, the reader learns that there is a significant effect of probability on the P3 wave but cannot tell whether the P3 was bigger for the probable stimuli or for the improbable stimuli. This is not an unusual example; I have seen Results sections like this many times in both journal submissions and published papers. Inferential statistics should be used as *support* for the waveforms and the mean values, and the statistics should never be given without a description of the actual data. Although this differs from how many papers are written, I am not alone in this view of how statistical analyses should be presented. For example, an editorial by Neil Macmillan in *Perception & Psychophysics* several years ago stated that "The point of Results sections, including the statistical analyses included there, is to make the outcome of the experiment clear to the reader...Readers should be directed first to the findings, then to their analysis" (Macmillan, 1999).

Some people object to this approach, claiming that a reader might be misled about an effect that was described but was not significant. However, if the Results section is written carefully, the reader will not be

misled. Also, this perspective fails to keep in mind something that we all learned in our introductory statistics courses: The fact that an effect is not significant does not mean that there was no real effect. If you are providing a thorough *description* of your findings, you should *describe* substantial differences between means whether the p value is .009 or .09 (but being careful to note that an effect with a p value of .09 was not statistically significant).

It is occasionally convenient to present the inferential statistics in one sentence and then immediately follow them with a description of the means or waveforms (rather than the other way around). There's nothing wrong with that, as long as the reader can easily understand how the statistics are related to the actual data.

### *Looking at the ERP Waveforms*

The main ERP-specific thing to consider carefully in the Results section is the figures showing the ERP waveforms. All ERP papers must provide figures showing the waveforms. This is one of the rules in the SPR publication guidelines, and it's an important rule. ERPs are simply too complex for the reader to be able to trust that all of the recording, processing, and analysis procedures were appropriate (and executed correctly) unless the paper provides an opportunity to see the waveforms. Also, as was described in Box 2.4 in Chapter 2, Kramer et al. (1985) showed that ERP experts do a good job of determining the underlying latent components when they see the actual waveforms (especially from multiple electrode sites), and presenting the waveforms is essential for this.

If difference waves are being analyzed, in most cases both the "parent" waveforms and the difference waves should be shown in figures. This is a rule in the SPR publication guidelines, and it should be followed in other journals as well in most cases. However, rules are made to be broken, and I consider it a good idea to break this rule when both of the following are true: a) a difference wave is being used in a very conventional way to isolate a component that is actually defined by a difference (e.g., N2pc, CDA, LRP), and b) the paper is in a journal with a broad readership that includes a lot of people who are not ERP experiments. When both of these are true, the advantages of seeing the parent waveforms are minimal, and keeping the number of waveforms to a minimum will help readers focus on the most relevant information.

As I've mentioned previously in this book, the first thing you should look at in a set of ERP waveforms is the prestimulus interval period. This gives you a good means of assessing the noise level in the data and will help you see if there is overlap or preparatory activity that might confound the poststimulus waveforms. You should also look at the first 100-200 ms after stimulus onset for signs of overlap, drift, or sensory confounds. In 99% of cases, you should not see any difference between waveforms prior to approximately 150 ms unless (a) there was a difference in the physical stimuli between conditions, or (b) the study involves shifts of spatial attention prior to stimulus onset. If neither of these is true, then there is almost always some kind of problem with the design or with the data (as discussed in several previous chapters). In addition, if the effects in the poststimulus waveform are no larger than the noise deflections in the prestimulus baseline period, you should be very skeptical about those effects (even if they're statistically significant). They're not necessarily bogus, but there is a good chance that they are.

### *How ERP Waveforms Should be Plotted*

Although it might seem trivial to plot ERP waveforms, I regularly read ERP papers in which the waveforms are plotted in a way that makes it difficult to perceive the key aspects of the waveforms. This section therefore provides some recommendations about plotting (see also the SPR publication guidelines).

First, it is almost always a good idea to show the data from multiple electrode sites, spanning the region of the scalp where the effects of interest were present. This information plays a key role in allowing experts to determine the underlying component structure of the waveform. However, including multiple electrode sites is not as important when components can be isolated by other means (e.g., by performing the subtraction procedure that isolates the lateralized readiness potential).

Second, plots of ERP waveforms should indicate the voltage scale and the time scale in a way that makes it easy for readers to assess amplitudes and latencies. The voltage scale should indicate whether positive or

negative is up (and if negative is up, I would recommend stating this in the figure caption as well). In addition, the electrode sites should be labeled in the figure (if a single site is shown, I often indicate the site in the caption as well as in the figure itself).

Figure 15.1 shows examples of a bad way to plot ERP waveforms (panel A) and a good way to plot ERP waveforms (panel B). The most egregious error in panel A is that there is no X axis line running through zero microvolts to provide an anchor point for the waveforms. It is absolutely essential to include an X axis line that shows the zero microvolt point and the time scale for each waveform (including a clear line at zero milliseconds). Without a well-labeled X axis line, the reader will find it difficult to determine the amplitude or the latency of the various points in the waveform. A separate scale marker, such as that shown in Figure 6.1A, is not enough, because it does not provide a visual reference point. For example, can you tell how much bigger the first component is in the right waveform than in the left waveform? And can you see that this component peaks earlier in the right waveform than in the left waveform?

The second most egregious error in Figure 15.1A is that the waveforms are not overlaid. This makes it very difficult for the reader to determine the exact pattern of effects. For example, can you tell exactly when the difference between the waveforms switches polarities? All key effects should be shown by overlapping the relevant waveforms. Figure 15.1A also fails to indicate time zero, and it is impossible to tell which polarity is plotted upward.

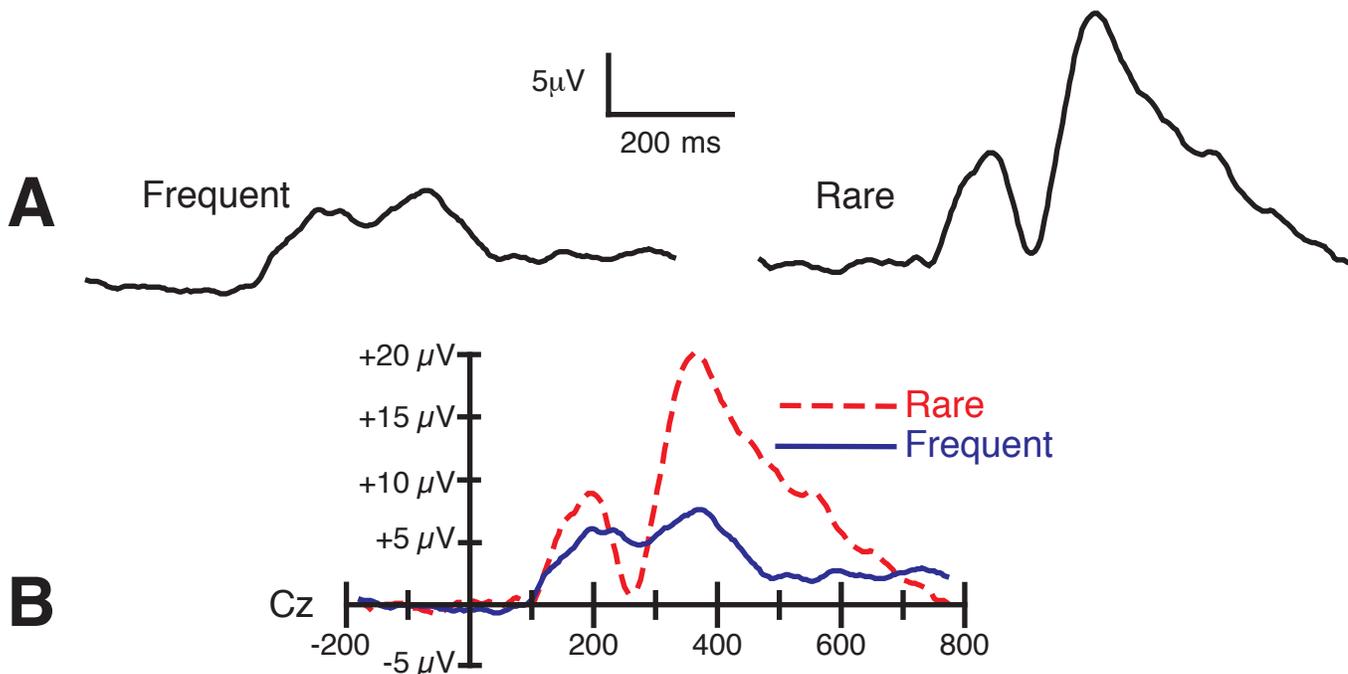


Figure 15.1. Examples of a poor format for plotting ERP waveforms (A) and a better format (B).

Panel B of Figure 15.1 solves these problems. It has an X axis running through zero microvolts with well labeled time ticks (including a line at zero milliseconds), and the waveforms are overlaid. The voltage calibration indicates the polarity of the waveform. The two waveforms are drawn in easy-to-distinguish line types, and a legend is provided in the figure to indicate what condition each line type represents (this is much better than providing this information in words in the figure caption or main text). The figure also indicates the electrode site at which the waveforms were measured.

It is also essential for the plot to show a significant prestimulus baseline period. As described in Chapter 8, I recommend a baseline period that is at least 20% of the overall epoch duration (e.g., you could use a 100-ms prestimulus baseline with a 500-ms epoch or a 400-ms prestimulus baseline with a 2000-ms epoch). Whatever baseline period was used for the baseline correction of the data should also be shown in the figure.

You should be careful when selecting the line types for the ERP waveforms. First, you need to use reasonably thick lines. Figures are usually reduced by a factor of 2 or 3 times for publication, making lines that were originally moderately thin become so thin that they are nearly invisible. And if someone makes a photocopy of the journal article, portions of the lines may disappear. Of course, you don't want to make the lines so thick that they obscure the experimental effects. When multiple waveforms are overlapped, the waveforms should be differentiated by both color and pattern (solid, dashed, dotted) so that they can be distinguished even in a black-and-white copy. You should typically avoid using different line thicknesses to differentiate the waveforms; this doesn't work as well, and the thin lines may disappear once the figure is reduced for publication. In almost all cases, you should avoid overlapping more than three waveforms (four if absolutely necessary). With more overlapping waveforms, the figure starts to look like spaghetti.

How many different electrode sites should you show? This depends, in part, on the expected audience. If the paper is to be published in an ERP-oriented journal, such as *Psychophysiology*, the readers will expect to see waveforms from a broad selection of electrode sites. If the paper is to be published in a journal where ERP studies are not terribly common, such as *Psychological Science* or *Journal of Memory & Language*, you should show only the key sites. If you have recorded from a large number of electrodes, it is pointless to include a figure showing the data from all the sites. If a figure contains more than 10-15 sites, the individual waveforms will be so small that they will provide very little information. It is much better to show a representative sample of the electrode sites and then use topographic maps to show the distribution of voltage over the scalp.

## The Discussion

I don't really have anything ERP-specific to say about the Discussion section. In general, a Discussion section should begin by reiterating the main results and describing the conclusions that can be drawn from those results. Then it is a good idea to describe possible alternative explanations and your arguments against them. This is typically followed by a discussion of broader implications of the findings (which I find to be completely useless in >90% of cases). In my view, if you've done a good job setting up the predictions in the Introduction and explaining the design and the data in the Methods and Results sections, the Discussion section should be brief and should serve mainly to bring all of the essential aspects of the study together in an easy-to-digest form.

## Some General Thoughts About Writing Papers

In this final section, I will provide a few general thoughts about writing journal articles, whether or not they include ERP data.

Good writing is an essential component of success in science. Even if you conduct the world's greatest experiments, you will have zero impact on the field if people don't read or can't understand your papers. Conversely, if you spend the time to write your papers well, your papers will be more likely to sail rapidly through the review process, get published in high-impact journals, and be read and cited by other researchers. Like anything else, writing is a skill that develops with practice. Some people find it easier than others, but everybody gets better with careful, thoughtful practice.

Box 15.2 describes the overall process that I use when I'm working with a student or postdoc who is taking the lead in writing a paper. If you have students or postdocs of your own, you might want to try it. If you're a student or a postdoc, you might suggest this approach to your mentor. I have found that it is both effective (in terms of creating a good final product) and efficient (in terms of the time required to create that final product).

The most important key to good writing is to see things from the perspective of the reader. What does the reader already know prior to reading your article? What does the reader want to know? In what order should the information be presented so that the reader can understand what you've done, what you've found, and what can be concluded? To answer these questions, you need to think about who will actually read your paper. For example, if you are writing the 30<sup>th</sup> paper on a given topic (with some slightly new twist) and publishing in the same journal where many of the previous 29 papers were published, most of the people who will read your paper will already be familiar with the general topic and won't need a longwinded introduction to that topic. If this topic has not been widely studied with ERPs, and your intended audience consists of people who are

experts in the broader topic but are not ERP experts, you should provide a lot of information about how ERPs will be useful in addressing this topic. However, if you're bringing this topic to an ERP-savvy audience in a journal that does not usually focus on this topic, you will need to provide a more detailed introduction to the topic.

*Box 15.2- My Approach to Writing with Students and Postdocs*

I have been extraordinarily fortunate over the years to have outstanding graduate students and postdocs, and they typically take the lead in writing journal articles. This box describes how I guide grad students and postdocs through the process of writing papers.

I like to begin the process of writing a journal article by having the student or postdoc put together a set of draft figures showing the central findings, along with all of the statistical analyses that support these findings. We then sit down with the figures and statistics and figure out what we've found, what we can conclude, and what additional analyses will be necessary. I find that this is much more efficient than starting with a complete Results section. If we start with a Results section, many major revisions are needed before we actually know what we've found.

Once we've agreed on the figures and stats, the student or postdoc writes the Method and Results sections. I then go through these sections, which typically leads me to suggest some additional analyses. After a few rounds of revisions, these sections are complete, and then it's time to write the Introduction. I then go through the Introduction in detail, and I also skim through the Method and Results sections to make sure they match the Introduction. After a few more rounds of revisions, it's finally time to write the Abstract and the Discussion.

I find that this process works much better than having the student or postdoc start by sending me a complete manuscript. If we start with a complete manuscript, 90% of it gets thrown out, which is an enormous waste of time and very disheartening.

One of the biggest mistakes to avoid is assuming that the readers of your paper are just like you, so think carefully about who will read your paper. And beyond that, try to put yourself in the reader's mindset so that you can anticipate what the reader needs to know and when the reader needs to know it.

Of course, you can't assume that all of your readers will be the same. In particular, some of the readers will just want to skim your paper to see what you found, whereas others (including the reviewers) will be reading it carefully and evaluating every single sentence and figure in the paper. I try to write papers so that they can be skimmed easily by a busy professor but also have all the information that will be needed by a graduate student who wants to use the same methods in a follow-up study. Here are some tips for making your paper easy to skim:

- Provide an overview of the task, including a figure, near the end of the Introduction. Make sure that you include a figure even if your task is so simple you think a figure shouldn't be necessary. Include a written overview of the task in the caption of the figure.
- Provide enough information in your data figures that someone can understand them without reading the Results section. This might even include a brief summary of the key results from a given figure in the caption for that figure.
- Recap the purpose of the study, the main results, and the main conclusions in the first 1-3 paragraphs of the Discussion section.

You should also keep in mind reader's memory limitations. A really good example of this is the use of abbreviations. Abbreviations are often convenient for the writer, but unfamiliar abbreviations are inconvenient for the reader. They require the reader to constantly engage in memory search, drawing processing resources away from the task of comprehending your text. So don't introduce new abbreviations if at all possible. In addition, if you write an important piece of information at one point in the manuscript (e.g., the Methods section), the reader may need to be reminded of this later in the manuscript (e.g., in the Results section).

It is also important to think about the paper from the perspective of the reviewers, because they are your first and most important audience. Reviewers are typically very busy people who spend considerable time reading and evaluating your manuscript without getting paid a cent, a yen, or a peso. It might not seem like it, but they are giving you a gift, and you should treat them accordingly. For example, make sure that your paper has been carefully proofread and is easy for them to read and understand. I have occasionally heard people say that they're not going to bother polishing the manuscript extensively because they're going to have to revise it eventually when they get the reviews. That's insulting to the reviewers, and it's likely to lead to a rejection without any opportunity to revise the manuscript. So you should carefully polish your manuscripts before submitting them.

When you get a set of reviews, think very carefully about what the reviewers are saying, and be very thoughtful and thorough in your response. It's very easy to think the reviewers are just being dumb (and occasionally they are), but it's more likely that you haven't explained something well or that there really is a flaw in your logic. So don't be dismissive of the criticisms. Instead, try to look at them from the reviewer's perspective and make sure that you are really addressing the criticisms that they raise. If you write a detailed and thoughtful response letter, bending over backward to accommodate the reviewers, you will have much more success getting your papers published.

My last comment in this section is something that I learned from Bruno Rossion. In the summer of 2013, I visited Bruno's lab in Louvain-la-Neuve, Belgium, and one evening we went to Brussels for a wonderful dinner. We started talking about the process of writing journal articles, and his lab manager asked me for general advice about writing. I answered by talking about the importance of understanding the audience and putting yourself in the mindset of the reader. Bruno followed up by saying that it's not enough to put yourself in the reader's mindset; you also need to *empathize* with the reader. Reading a journal article is hard work, and you need to feel the reader's pain and joy. This will lead you to devote the effort required to minimize the pain and maximize the joy, which will make your papers a pleasure to read. And if your papers are a pleasure to read, your research will have a much larger impact on the field.

### **Top Ten Reasons for Rejecting an ERP Manuscript**

Over the years, I have served as a reviewer or action editor for hundreds of ERP manuscripts that have been submitted for publication. The following is a list of the most frequent ERP-specific problems that I have seen (in no particular order). You might want to use this as a list of things to check when you're reading or reviewing ERP papers. Some of these problems can be easily fixed in a revision, but others require new experiments to address.

- 1) Baseline problems, including differential overlap or preparatory activity between groups or conditions (see Chapter 11) and effects that begin so early that they are clearly artifactual (see Figure 4.4 in Chapter 4).
- 2) Physical stimulus confounds (violations of the Hillyard Principle, as described in Chapter 4). These are sometimes obvious confounds (different stimuli used in different conditions), but they are sometimes less obvious (e.g., differential adaptation as a result of different sequences of stimuli).
- 3) Use of a biased measure of amplitude or latency in conditions or groups that have different noise levels (see online supplement to Chapter 9).
- 4) Inappropriate filtering, such as using an extreme high-pass filter or a narrow bandpass filter that might be causing substantial distortion of the waveforms.
- 5) Failure to provide clear and specific predictions about the expected pattern of ERP results.
- 6) Effects of blinks or eye movements, including direct effects (i.e., electrical potentials arising from the artifacts that contaminate the ERPs) and indirect effects (changes in the sensory input).

- 7) Statistical problems, especially inflation of the Type I error rate as a result of using the observed waveforms in selecting the time windows or electrode sites used in the statistical analyses (see Chapter 10).
- 8) Failure to isolate the component of interest (see Chapter 2, especially Figure 2.5). For example, an effect from 400-600 ms is interpreted as a change in the P3 wave even though it might have been some other component that changed. Similarly, a large change in the amplitude of one component might cause a modest shift in the latency of a temporally adjacent component.
- 9) Differences in state across groups or conditions. For example, if one condition is more difficult than another, this might lead to increased arousal, which could in turn lead to changes in the amplitudes of both early and late components (see, e.g., Figure 3.2 in Chapter 3).
- 10) Reliance on source localization methods to draw major conclusions without careful consideration of the margin of error of the localization. That is, how confident can we be that the localization is correct, and would the major conclusions differ if the localization was incorrect?

## References

- Bem, D.J. (1987). Writing the empirical journal article. In M. P. Zanna & J. M. Darley (Eds.), *The Compleat Academic: A Practical Guide for the Beginning Social Scientist* (pp. 171-201). New York: Random House.
- Keil, A., Debener, S., Gratton, G., Junhöfer, M., Kappenman, E.S., Luck, S. J., Luu, P., Miller, G., & Yee, C.M. (2014). Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography *Psychophysiology*, 51, 1-21.
- Kramer, A.F. (1985). The interpretation of the component structure of event-related brain potentials: An analysis of expert judgments. *Psychophysiology*, 22, 334-344.
- Macmillan, N.A. (1999). Editorial. *Perception & Psychophysics*, 61, 1-2.
- Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Jr., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, 37, 127-152.
- Urbach, T.P., & Kutas, M. (2002). The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology*, 39, 791-808.
- Urbach, T.P., & Kutas, M. (2006). Interpreting event-related brain potential (ERP) distributions: implications of baseline potentials and variability with application to amplitude normalization by vector scaling. *Biological Psychology*, 72, 333-343. doi: 10.1016/j.biopsycho.2005.11.012