Chapter 6 Supplement

A Closer Look at ICA-Based Artifact Correction

In this supplement, I will describe some of the practicalities of ICA-based correction, using data provided by Carly Leonard. I will begin with a more detailed general description of how ICA-based correction works, and then I will show you what the process looks like when applied to real data.

The Essence of ICA

To explain how ICA works, I first want to remind you about how sources of voltage in the brain mix together when they are conducted to the electrodes (see the section on "The Forward Problem and the Superposition of Components on the Scalp" in chapter 2). As shown in figure S6.1, each component in the brain has a source waveform that describes the time course of activity for that component over some period of time (whether on single trials or after averaging). The waveform recorded at a given scalp site is a weighted sum of the source waveforms. The weights depend on the position and orientation of each source relative to each electrode site, along with the conductivity of the tissues. For each source, there is a weight for each electrode site, and these weights define the scalp distribution of that source. If we have S sources and E electrode sites, we will have an S × E matrix of weights. This *mixing matrix* defines how the different sources mix together as the voltages flow through the brain, skull, and scalp. Artifacts such as blinks and eye movements—along with EMG, line noise, and so forth—are simply additional sources that can be included in this S × E matrix.

The goal of ICA is to find an *unmixing matrix* that will allow us to take the observed waveforms from the scalp electrodes and calculate the time courses of the underlying components. This unmixing matrix is the inverse of the mixing matrix. However, unless we know the locations of the sources and all of the conductivities, we don't know the values in either the mixing matrix or the unmixing matrix. As will be discussed in online chapter 14, source localization techniques use the biophysics of voltage conduction to estimate the mixing matrix and the unmixing matrix. In contrast, ICA uses the statistical properties of the observed EEG data to create an unmixing matrix. ICA does not "know" anything about the biophysics of EEG. For example, it knows nothing about where each channel is located, and it doesn't even know that the data represent consecutive time samples. Thus, ICA finds statistical abstractions of components, not neuroanatomic sources, and the ICA components may or may not correspond to true neural sources or true

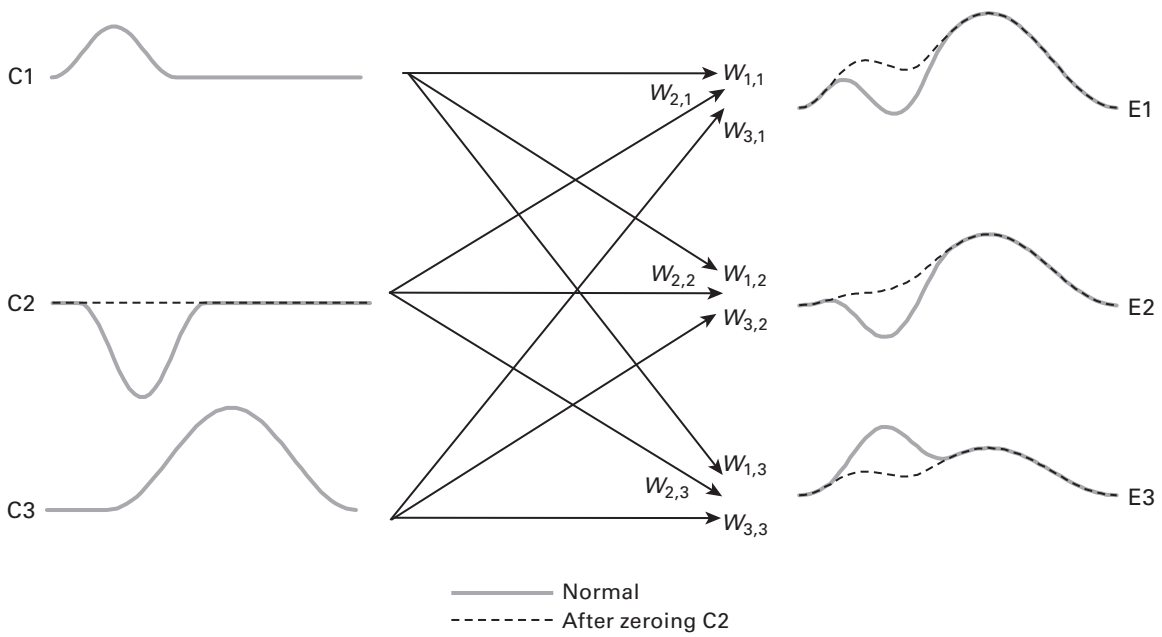

Figure S6.1
Reminder of how intracranial ERP sources propagate to the scalp electrodes. The voltage at each electrode is simply the weighted sum of each internal source. If one source is an artifact, we could eliminate the contribution of this source by setting it to zero and then recomputing the voltage at each electrode site. This is illustrated here by showing what happens when we set the activity of component C2 to zero and recompute the waveforms at each electrode site (blue line).

components (as defined in chapter 2). To distinguish between how the term *component* is used by ICA and how it is used in other contexts, I will use *IC* (independent component) to refer to the ICA components.

I won't provide the details of how ICA comes up with an unmixing matrix (for an excellent review, see Makeig & Onton, 2012), but here's the gist. The raw EEG and EOG data (the time series of voltage values for each channel) are fed into a neural network, and this network uses a learning algorithm to generate an unmixing matrix that leads to ICs that are *maximally independent*. The number of ICs will be equal to the number of channels of EEG/EOG data. For each time point in the EEG/EOG data, each component will have a magnitude, which indicates the strength of that component at that time point.

Figure S6.2 shows this in a flow diagram. The first step is that a neural network processes the data to find the unmixing matrix. When the unmixing matrix is multiplied by the matrix of EEG data, this gives us the time course of each IC. The unmixing matrix is labeled W in figure S6.2, and it's a two-dimensional matrix of electrodes and ICs. It provides a set of weighting values that "pull out" the time course of the ICs from the EEG matrix (which is a two-dimensional matrix of electrodes and time points). Thus, the IC activation waveforms shown in figure S6.2 are analogous to the source waveforms in figure S6.1. And just as the voltage at each electrode

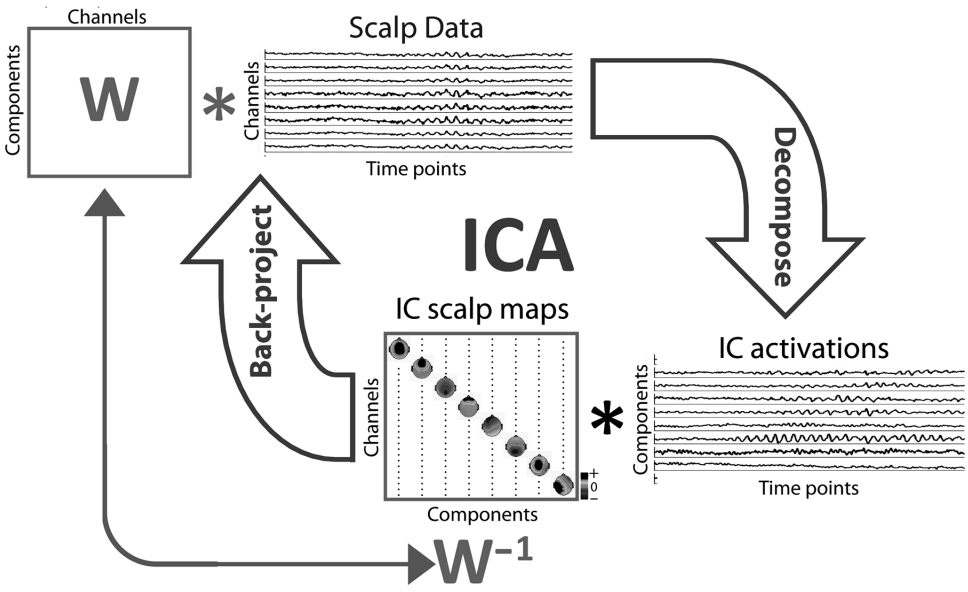

Figure S6.2
Flow diagram of ICA. The scalp EEG/EOG data are passed through a neural network to obtain the unmixing matrix, W. If the unmixing matrix is multiplied by the EEG/EOG data at each time point (the scalp data), this yields the activation of each of the independent components (ICs) at each time point (the IC activations). The mixing matrix, W^{-1}, is the inverse of the unmixing matrix, and it represents the scalp distribution of each IC. If we multiply the IC activations by the mixing matrix, we recover the original scalp data. Reprinted with permission from Makeig and Onton (2012). Copyright 2012 Oxford University Press.

site is a weighted sum of all source waveforms in figure S6.1, the voltage at each electrode site in figure S6.2 is the weighted sum of all the IC activations, where the weights are given by the mixing matrix (which is the inverse of the unmixing matrix and is labeled W^{-1} in figure S6.2). In other words, you can exactly re-create the original EEG/EOG data by passing the IC activations through the mixing matrix.

You can also think of the mixing matrix as consisting of the scalp distributions of the ICs. For every microvolt of activity in a given IC, the scalp distribution of the IC indicates the number of microvolts you will get at each electrode site. One caveat for ICA, however, is that the units of activation for each IC are arbitrary. Thus, you can have a huge activation for a given IC simply by having small weights between that IC and each electrode site in the mixing matrix. Similarly, a positive voltage in the electrodes can be a result of a negative value for the activation combined with a negative value for the weights. Thus, the absolute magnitudes of the activation values are informative only when considered along with the weights (scalp distributions).

Now that you have a sense of the sequence of operations, it's time to consider what it means that ICA extracts components that are *maximally independent*. This basically means that knowing the strength of one component at a given moment in time provides no information about the strength of another component at that time. In other words, if you were to plot the activation of two different

ICs over the course of a session, you would find no relationship between the activation of one IC and the activation of the other IC. For example, figure S6.3 shows a scatterplot of two ICs, showing how the activation of one IC is related to the activation of the other IC across a set of time samples. The scatterplot is just an unstructured cloud of points, with no relationship between the two components (whether linear or nonlinear). This is what it means for the components to be independent. Although this rule for determining the ICs is not based on dipoles and conductivities, it often comes up with ICs that are consistent with what we know about EEG (e.g., the ICs often have a bipolar scalp distribution; see Delorme, Palmer, Onton, Makeig, & Oostenveld, 2012).

The ability of ICA to accurately recover the underlying component structure of the data will depend on the extent to which the actual underlying components are statistically independent. People sometimes argue that ICA is an absurd approach with EEG data because every part of the brain is ultimately connected to every other part of the brain, and therefore nothing is independent. However, ICA requires only that the strength of one IC cannot be predicted by the strength of the other ICs *at each individual time point*. If one IC reliably follows another by 100 ms, these two ICs might still be independent at each individual time point. For example, even if a subject always blinks immediately after responding, the response-related activity will not occur at the same time as the blink-related activity, and this may lead to statistically independent components for the response and the blink. Moreover, ICA may be quite accurate if the underlying components are mostly independent, even if they are not completely independent. In many (but not all) cases, artifacts would be expected to be largely independent of brain activity (in this specific sense of the term *independent*), and ICA is therefore well suited for artifact correction.

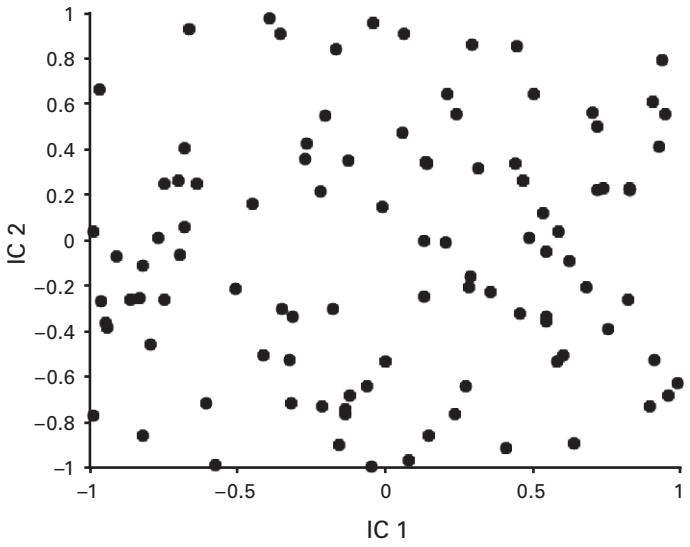

Figure S6.3
Example of the independence of two independent components (ICs). Each point in the scatterplot represents the activation of IC 1 (*X* axis) and IC 2 (*Y* axis) at a given point in time. There is no consistent relationship between IC 1 and IC 2, and this is what it means to say that they are statistically independent.

ICA and Artifact Correction

Figure S6.4 shows how ICA is applied in artifact correction (from the study of Jung et al., 2000). A 3-s period of EEG is shown at the left. A blink is clearly visible at the frontal electrodes in the original EEG data. EMG is also visible at the T3 and T4 electrodes (which correspond to T7 and

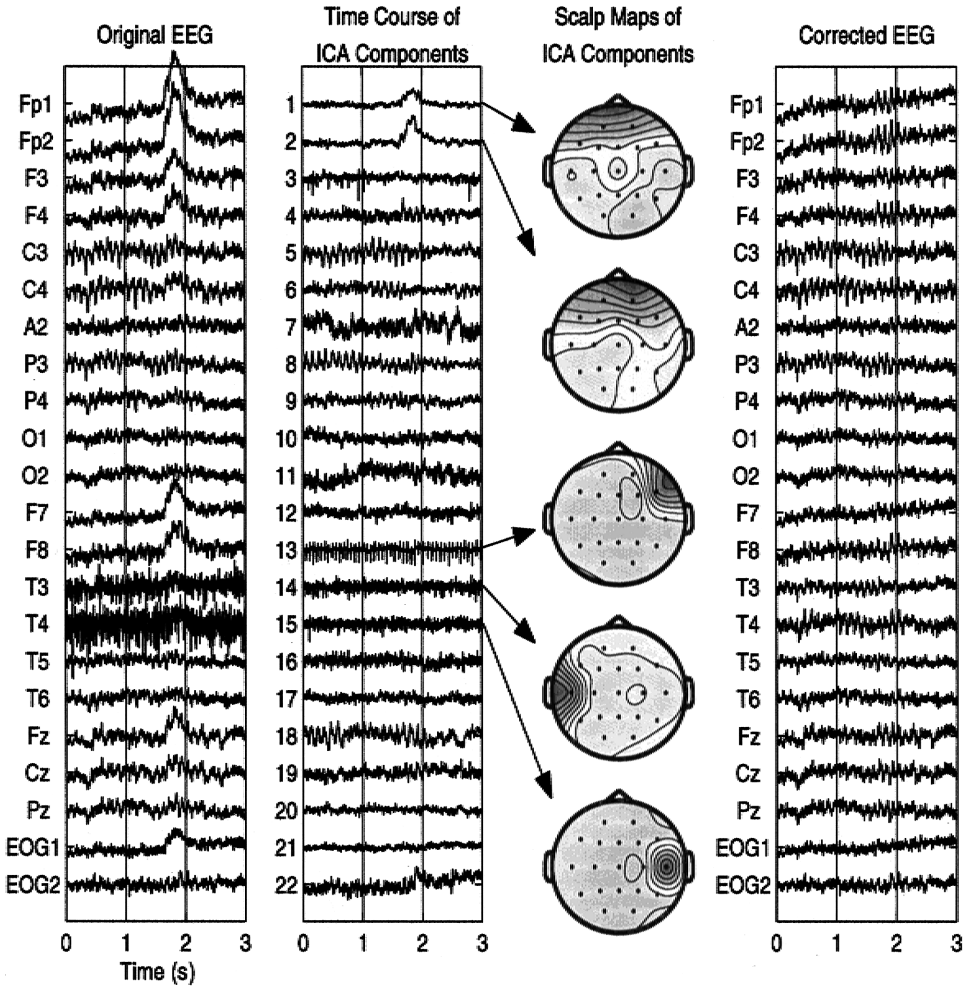

Figure S6.4
Example of the use of ICA in artifact correction. The leftmost column shows the raw EEG/EOG data from a 3-s time period. The next column shows the time course of activation for each of the ICA components. The next column shows the scalp distributions of several of the ICA components. The rightmost column shows the reconstructed EEG data, computed by passing the ICA activations through the mixing matrix but without including the artifactual components. Reprinted with permission from Jung et al. (2000). Copyright 2000 Society for Psychophysiological Research.

T8 in the current electrode naming scheme). In addition, EMG activity was also present at right frontal electrodes (F4 and F8, but this is difficult to see in the figure).

The activation time courses of the ICs from this 3-s period are also shown (note that the EEG from the whole session was used to derive the unmixing matrix, not just this short time period). IC 31 and IC 32 have approximately the same time course as the blink, with a peak at approximately 1.8 s. The scalp distributions of these ICs have the distribution that you would expect for a blink, with the strongest weights near the front of the head. In most cases, a blink will be confined to a single IC. In some cases, like this one, it appears in two ICs. This could reflect either some independence between the two eyes or some independence of the movement of the eyelid and the rotation of the eyes.

IC 14 and IC 16 correspond to the muscle noise in the T3 and T4 electrode sites. The millisecond-by-millisecond spiking of these muscles would be expected to be independent, so they appear as separate components. If you could see an expanded view of the data, you would see that the time course of the spikes in IC 14 corresponds with the time course of the EMG spikes in T3, whereas the time course of the spikes in IC 15 corresponds with the EMG spikes in T4.

The contribution of a given IC to the observed EEG over a segment of time is determined by multiplying the time course of the IC by its scalp distribution (i.e., the product of the IC and its scalp distribution). The overall EEG can be reconstructed by taking the sequence of values resulting from the produce of each IC and its scalp distribution and simply summing them. We can therefore eliminate the contribution of the artifacts by just zeroing the time course for the corresponding ICs before we reconstruct the EEG. This is illustrated in figure S6.1, which shows how the voltage at each electrode site is simply a weighted sum of the underlying components (which is true whether we have the "real" components or are using estimates of these components from ICA). If we set the time course of one component to zero, this component will be missing from the scalp electrodes. For the real data shown in figure S6.4, the right side shows the reconstructed EEG after zeroing the time courses of ICs 1, 2, 13, 14, and 15. You can see that the blink has been eliminated at the frontal sites and that the EMG noise has been eliminated from T3 and T4.

A More Detailed Example

In this section, I will go through some of the details of an experiment that Carly Leonard analyzed. In this experiment, Carly recorded from 37 electrodes, including 32 scalp sites, left and right horizontal EOG sites, left and right mastoids, and a vertical EOG site under the left eye. Using EEGLAB Toolbox, she applied ICA to the continuous EEG data from each subject to get the ICs for each subject (note that ICA is always done separately for each subject). The data were high-pass filtered with a half-amplitude cutoff of 0.05 Hz prior to the ICA.

Figure S6.5 shows the scalp maps of the ICs she obtained for one of the subjects. ICA is done on the individual subjects, because the scalp distribution of an artifact (and the other components) will not be exactly the same in each subject. In ICA (and related methods), the number of ICs is always equal to the number of electrodes. This constraint is necessary if you want to be able to perfectly reconstruct the original EEG from the IC time-course waveforms and scalp distributions. Because it uses a neural network, ICA extracts ICs in a random order. However, software packages typically present them ordered according to the amount of moment-by-moment variance in EEG/EOG amplitude they explain, and that is how they are ordered in figure S6.5.

Once you've extracted the components, the next step is to determine which components reflect artifacts. Carly did this manually by looking at the scalp distributions and time courses of the ICs to determine which ones reflected blinks and eye movements. Tools now exist that will do this automatically, but you should check the results by eye to make sure that everything is working correctly (at least when you first start using ICA).

In the example shown in figure S6.5, IC 2 had a scalp distribution that would be typical for an eyeblink, and IC 4 had a scalp distribution that would be typical for a horizontal eye movement.

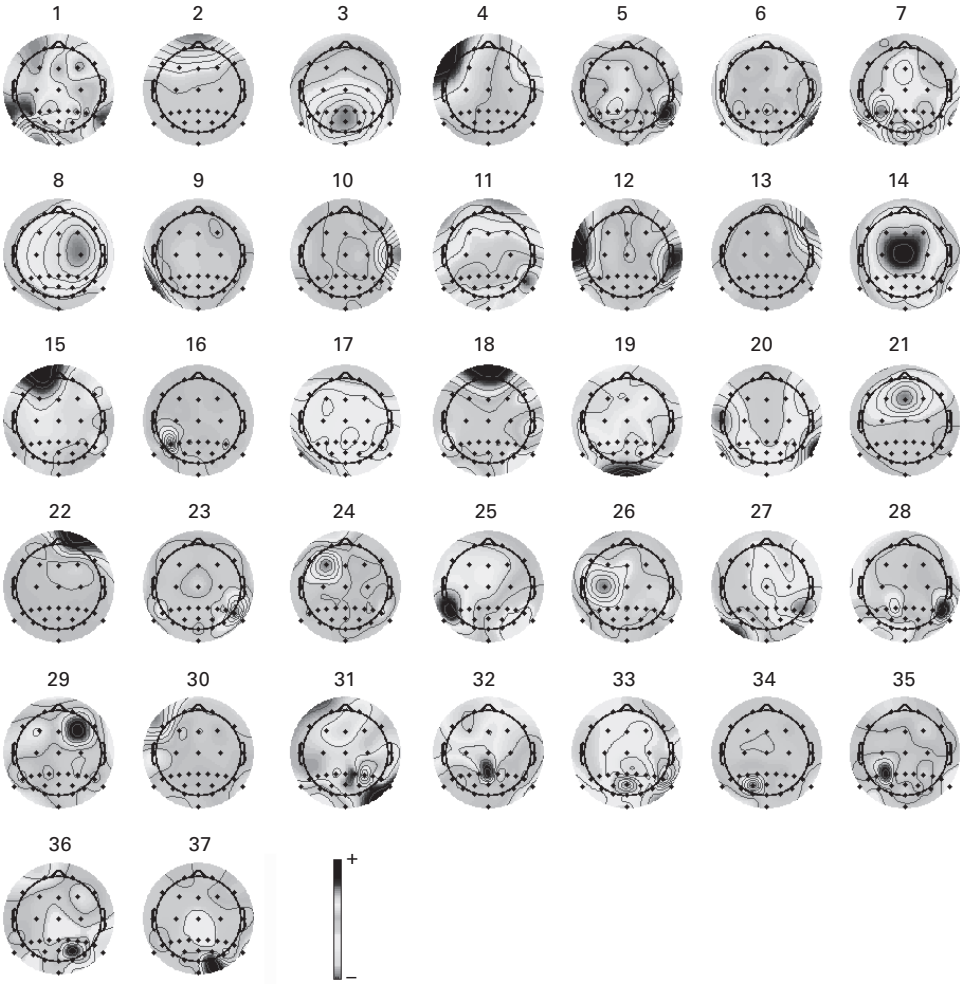

Figure S6.5
Maps of the 37 independent components from a single subject in the example study. Component 2 represents blinks, and component 4 represents horizontal eye movements.

To verify that these ICs reflected blinks and eye movements, respectively, Carly also determined whether their time courses matched the time courses of the blink and eye movement artifacts in the EOG channels. A segment of EEG/EOG data with a blink is shown in figure S6.6, along with the time course of each IC for the same time period (to make things easier to see, only seven of the 37 EEG channels are shown, and only the first seven ICs are shown). You can see that the blink is opposite in polarity at the VEOG site relative to the Fz site. Some of the blink activity can also be seen at the HEOG sites, both of which are referenced to a mastoid electrode (i.e., this is not a bipolar HEOG channel). If the HEOG electrodes are slightly above or slightly below the midline of the eyes, the eyeblink artifact will be visible at these sites. This eyeblink is also clearly present in the IC time course for component 2 (figure S6.6B), with the same shape and timing as in the corresponding EEG segment. Thus, this component has the appropriate scalp distribution for a blink (shown in figure S6.5) and shows a blink-like deflection in its time course whenever a blink occurred in the EEG/EOG recordings. Nothing blink-like was visible in any of the other components during this time period. Thus, we can eliminate the blink activity by zeroing component 2.

The original and corrected EEG/EOG are shown in figure S6.6C. There is essentially no difference between the corrected and uncorrected data during most of the epoch (the two lines overlap so precisely that it is difficult to see that two lines are actually present). During the time period of the blink, however, most of the blink-related activity has been removed from the corrected data. A close-up is shown in figure S6.6D, showing that some spiky activity at the beginning of the blink was not removed. This is probably EMG from the muscles that produce the blink, which does not have the same scalp distribution as the main blink and was presumably present in a separate IC.

Figure S6.7 shows comparable data from a time period that contained two eye movements. These eye movements appear as sudden voltage deflections in the three EOG channels. Corresponding deflections can be seen in the time course of component 4, with exactly the same timing as the deflections in the EOG channels. This correspondence, combined with the scalp distribution shown in figure S6.5, indicates that this component reflects horizontal eye movements.

When correction was performed (figure S6.7C), the corrected and uncorrected data matched extremely well except for the voltages caused by the eye movements. This tells us that the correction worked reasonably well.

Eye movement correction did not work perfectly for all subjects in this experiment. For many subjects, eye movement activity could be seen in the time courses of several components. This may reflect the fact that subjects in this experiment made eye movements of many directions (not just horizontal and vertical). The scalp distribution for a diagonal eye movement is the weighted sum of the scalp distributions for pure horizontal and pure vertical eye movements, so it is possible in principle to represent any saccadic eye movement with two underlying components. However, the time course of the horizontal and vertical components will be very similar for diagonal eye movements, and ICA requires that the time courses of all ICs are independent of each other. This may make it impossible to create separate ICs for the horizontal and vertical components when the angle of the eye movements vary considerably from trial to trial. Consequently, I urge you to be very cautious if you use ICA to correct for eye movements.

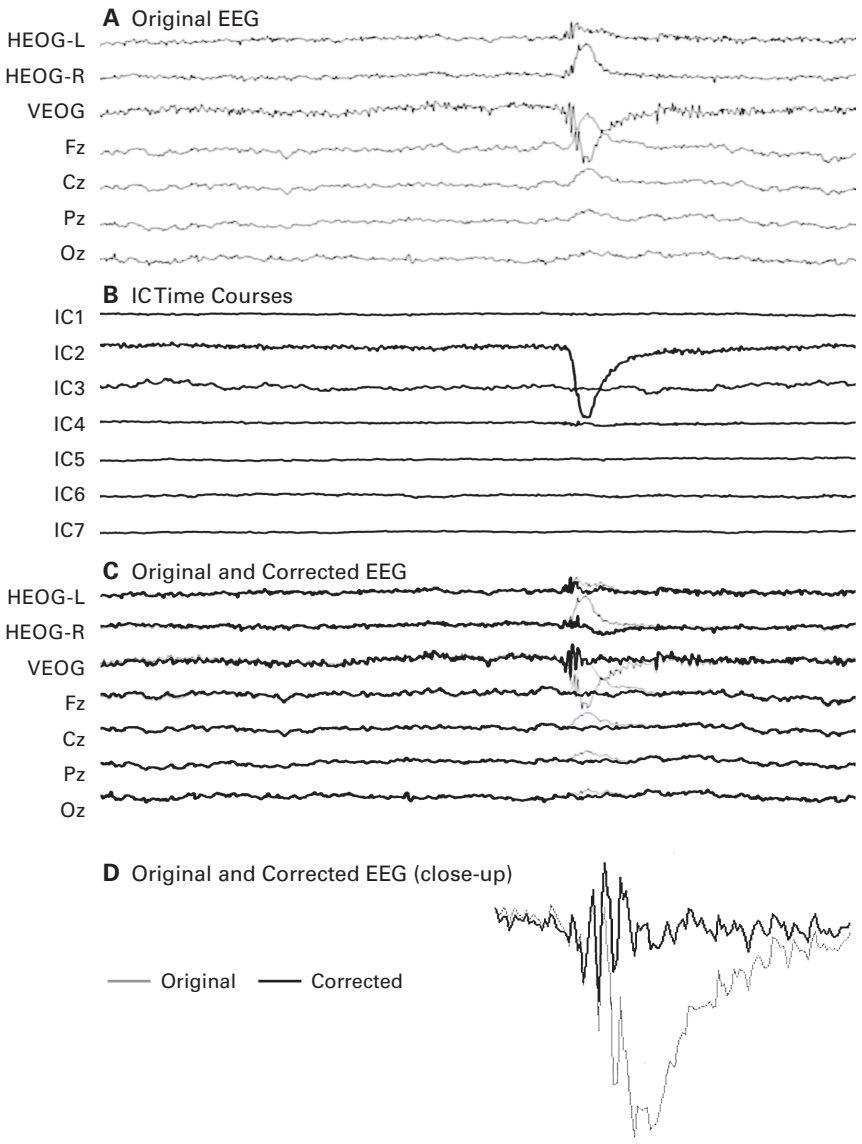

Figure S6.6
(A) Segment of the original EEG from the example study, showing an eyeblink. (B) Activation time courses of the first seven independent components from the same time period. Note that a blink-like deflection is present in component 2 at exactly the same time as the blink in the original EEG. (C) Original EEG data overlaid with the corrected data (reflecting the removal of component 2 and component 4). (D) Close-up view of the corrected and uncorrected data at the time of a blink.

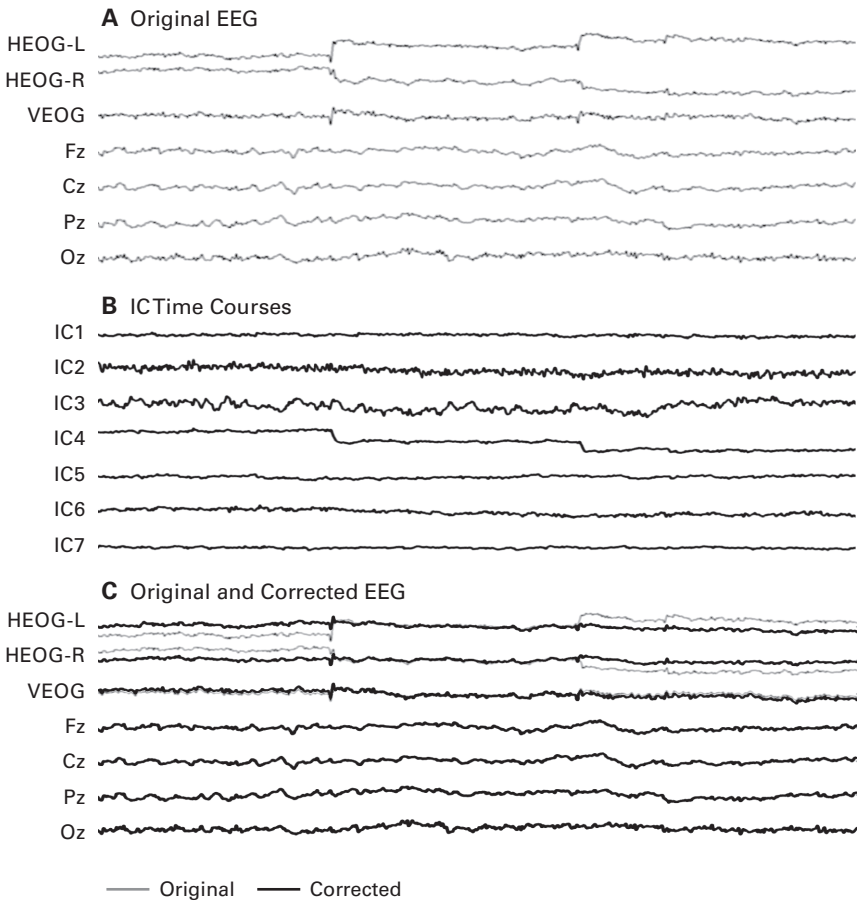

Figure S6.7
(A) Segment of the original EEG from the example study, showing saccadic eye movements. (B) Activation time courses of the first seven independent components from the same time period. Note that voltage transitions in component 4 occur at the same time as the saccade-related voltage transitions in the HEOG data in panel A. (C) Original EEG data overlaid with the corrected data (reflecting the removal of component 2 and component 4).

Optimizing ICA-Based Correction

Because the number of ICs is always equal to the number of channels, ICA will inevitably collapse multiple true components into a single IC and/or divide a single true component among multiple ICs. These problems are much less likely to create significant distortions for underlying components that account for large portions of the moment-to-moment variance in voltage. Consequently, ICA-based correction will work best if the artifacts being removed account for a large portion of the variance. Large artifacts, such as blinks and large eye movements, obviously

tend to account for a large portion of the variance, but only if they occur relatively frequently. If an artifact is rare, you might have better luck simply excluding the trials that contain that artifact rather than trying to remove the artifact with ICA.

Another implication of this is that your correction will work better if you can eliminate other large sources of variance. There are two main ways to do this. First, ICA experts recommend that you delete sections of data that have very large noise (C.R.A.P.), such as periods of time between trial blocks when the subject is moving, talking, and so forth. An example is shown in figure S6.8. You might be tempted to eliminate every period where you see a burst of EMG noise, but that would be going too far. Just delete sections with really crazy-looking EEG/EOG data.

A second large source of noise is the offsets and gradual drifts that were described in chapter 5. If you apply correction to epoched EEG segments rather than the continuous EEG, you can eliminate much of the offset by means of baseline correction. However, as will be discussed in chapter 8 and online chapter 11, baseline correction sets the scalp distribution to zero during the prestimulus period, and this can distort the poststimulus scalp distribution (see figure 11.4). ICA assumes that the scalp distribution of a component is exactly the same at every time point of every trial, so anything that distorts the scalp distribution can lead to poor extraction of ICs.

When my lab first started using ICA for artifact correction, we were advised by Arno Delorme (one of the EEGLAB developers) to apply it to the continuous EEG/EOG data to avoid this problem, and we have found this to work well. A high-pass filter with a cutoff somewhere between 0.01 and 0.01 Hz can be applied to minimize offsets and drifts. A downside with this approach is that it is somewhat tricky to correct ocular artifacts throughout the data and also reject trials that contain these artifacts at the time of the stimulus. That is, if you apply artifact correction to the continuous data to remove the blinks and eye movements, and then you epoch the data to perform artifact rejection, you can no longer determine whether an artifact is present in the epoched data. We spent quite a bit of time developing a procedure that makes this possible in ERPLAB Toolbox.

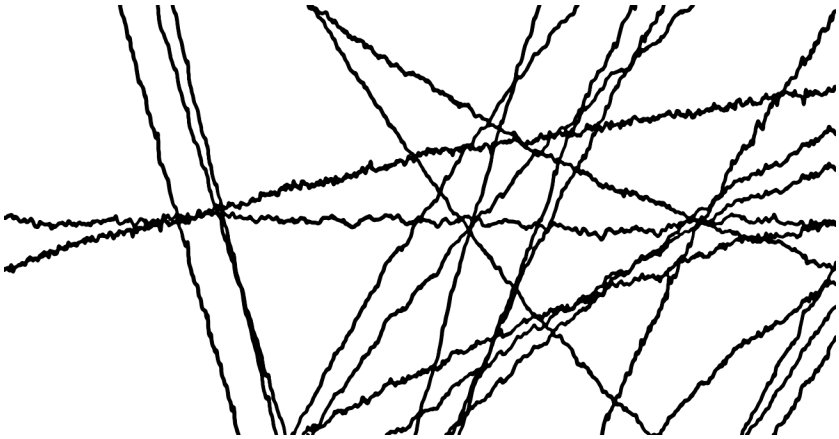

Figure S6.8
Example of the type of "crazy" EEG segment that should be eliminated prior to ICA-based artifact correction

I have heard many people tell me that ICA-based rejection works well for them when applied to epoched EEG segments, but there is not much systematic evidence about the best approach. One study found that the reliability of the ICA components (i.e., the similarity of the ICs when applied to different halves of the same set of data) was substantially better when the whole epoch was used for baseline correction rather than a 100-ms prestimulus interval (Groppe, Makeig, & Kutas, 2009). However, they did not test a longer prestimulus interval, and they did not test the effects of first applying a high-pass filter. In addition, they focused on reliability, which is not the same thing as accuracy. Thus, this study was very useful but was only a first step. There is a real need for good simulation studies that examine these issues, exploring both the accuracy and the reliability of the correction.

Given that these practical issues have not yet been resolved, what should you do? First, by the time you are reading this, someone may have conducted a definitive study, so I would recommend conducting a literature search before you start using ICA-based correction. If this has not yet been resolved, I would recommend trying ICA-based correction several ways and seeing what works best with your data. You can informally test the reliability of the results (e.g., by applying correction to the first half and second half of each session and comparing the results). You can also see whether the correction is changing the waveforms only at the time of an artifact (as in figures S6.6 and S6.7). My guess is that applying ICA to the continuous data with a high-pass filter at 0.01–0.1 Hz will work best, but that is just a guess. Whatever you do, you should report it completely in your methods sections.