Chapter 3 Supplement

Linking ERPs with Neural and Cognitive Processes

In this supplement, I will consider what it really means to link an ERP component with a specific function, the challenges involved in creating such a link, and the approach that I would advocate for future research on these links.

I'd like to warn you that I am rather pessimistic about the possibility of determining, with high levels of certainty, the link between a given ERP component and a highly specific functional process. But this pessimism applies to linking any physiological measure—including the fMRI BOLD signal and even single-unit activity—to a specific functional process. Creating these links is more complicated than you might think. However, ERPs and other physiological measures are still incredibly useful for answering fundamental questions about functional processes. The key is to go beyond simply asking whether a given neural measure is present under a given set of conditions. Chapter 4 presents some specific strategies for avoiding the need to identify specific components and their links to specific functional processes.

If you are interested in reading more about this topic, I recommend Manny Donchin's review papers, particularly Donchin et al. (1978) and Donchin and Coles (1988). Manny has always been one of the most insightful thinkers and clearest writers about the general nature of ERPs and ERP components. You might also want to read the chapter that Emily Kappenman and I wrote for the *Oxford Handbook of ERP Components* (Kappenman & Luck, 2012).

Lists of Antecedents versus Functional Theories

In discussing the functional significance of the P3 wave, Donchin and Coles (1988) noted that a theory of the process that an ERP component represents involves more than describing the necessary antecedents for the component (i.e., the conditions that must be present for the component to occur). As an analogy, they proposed a thought experiment in which you hear a whirring sound coming from your computer at certain times and want to know what underlying process this whirring sound represents. You might find that the sound occurs only when the computer is on, only when an external hard drive is attached, and only when the hard drive is turned on.[1] You might also find that the sound reliably occurs for several seconds when the computer is restarted, that it also happens whenever certain applications are launched, that it frequently but not always happens when you execute a *save* command, and that the sound is localized to the external hard drive. These might be highly reliable, statistically significant observations about the antecedent conditions for hearing the whirring sound, but this list of antecedents does not constitute a theory of the functional significance of the whirring sound. That is, although they are clues, the antecedents alone do not tell what functional process is occurring when the whirring sound occurs. Similarly, a list of the antecedents for the occurrence of a P3 wave—the eliciting stimulus category must be rare, the subject must be paying attention, the stimulus must be discriminable—is not the same as having a theory of the process that the P3 reflects (the computational process served by the circuit that generates the P3). To count as a functional theory, a hypothesis about a component must specify what is accomplished by the brain activity that generates the component. That is, a functional theory must specify what the brain activity is "for."

If we do enough experimentation on the whirring noise coming from the computer, we might develop a very detailed and powerful theory of the antecedents of the noise, stating that it occurs whenever data must be transferred between the main computer and the hard drive. That would be a major step forward and go way beyond the original list of antecedent conditions. We might even propose a functional theory of the whirring noise, saying that it is generated by the process of transferring information between the computer and the hard drive. This would be a true functional theory because transfer of information has a clear purpose in information processing. With this functional theory of the whirring noise, we could try to make precise, quantitative predictions about how the whirring noise varies as a function of what we are doing on the computer. However, this theory would be incorrect. That is, it would not be correct to say that the whirring noise is created by the actual process of moving information between the main computer and the hard drive. Instead, the whirring noise actually comes from the hard drive's motor, which is typically activated when information is transferred to and from the hard drive. That is, the function of the process reflected by the whirring noise is to move a platter that contains magnetic fields that encode bits of information. Similarly, even if the P3 is closely associated with context updating, the neural activity involved in context updating might not directly generate the P3. Instead, the P3 might be generated by a neural "housekeeping" activity that is needed whenever context updating occurs but is not an intrinsic part of context updating. Along these lines, Donchin and Coles (1988) noted that their theory states that the P3 "is a manifestation of a process invoked in the service of the updating process, not necessarily the updating per se" (p. 357).

Notably, this analogy creates both a "high bar" for any theory of the functional significance of the P3 wave (because it must be more than a list of antecedent conditions) and a lot of "wriggle room" for explaining away results that don't fit with a given theory (because one could always propose that the component represents some ancillary process that is imperfectly associated with the process of interest). This is not necessarily a major problem because the history of science shows us that major theories almost always need a little wriggle room—especially when they are new—to avoid prematurely rejecting them before we fully understand the broader theoretical domain. However, it illustrates some of the challenges involved in relating a voltage recorded on the scalp to a specific psychological or neural function.

If you think back to the many ERP components that you've read about in this chapter, you may realize that we do not have functional theories for many of them. We mostly have lists of antecedent conditions. The N170, for example, seems to occur whenever a stimulus is processed as a face, but this does not tell us what the N170 process actually does. Does it calculate relative distances between major facial features? Does it attempt to link a face with a long-term memory representation of an individual? We don't really know. Similarly, the MMN appears to occur when a current stimulus is compared with a memory of a previous stimuli, but what is the purpose of this comparison? The amplitude of the CDA is related to the number of items that are active in visual working memory, but does the CDA reflect the processes involved in keeping the items active or, as proposed earlier, the processes involved in keeping the representations from interfering with each other?

My initial theory of the N2pc would seem to count as a functional theory. That is, I proposed that it reflects a process that suppresses distractors so that they would not lead to ambiguous neural coding when multiple items simultaneously appear within a given neuron's receptive field (Luck & Hillyard, 1994b; Luck, Girelli, McDermott, & Ford, 1997). However, this theory appears to be incorrect, and the revised version I described earlier in this chapter is not really a functional theory.

The context updating theory of the P3 wave is, of course, a functional theory. But the evidence in favor of it is not particularly strong, and it's not very specific (especially if the P3 does not reflect the updating process, per se, but some other process that is invoked in the service of the updating process). Peter Hagoort's theory that the N400 reflects the process of integrating a new stimulus into the current semantic context representation would seem to count as a functional theory, but it's not terribly specific. That is, I'm not sure exactly what is involved in this integration process. There are some very specific functional theories of the ERN and anterior N2, as described earlier in the chapter, but the field has not yet converged on which theory is correct.

This is a little depressing, because almost 50 years have passed since the discovery of the first cognitive ERP component (Walter, Cooper, Aldridge, McCallum, & Winter, 1964), and we still don't have any really solid, highly specific functional theories for any of the major cognitive ERP components. The next section will describe some reasons for this failure, and the section after that will attempt to cheer you up by explaining why this doesn't matter as much as you might think. As a bit of foreshadowing, consider the whirring noise analogy: If we hypothesized that it reflected information transfer between the computer and the external hard drive, which is not quite right, we would still be able to use the whirring noise to learn many interesting facts about how the computer worked.

Why Functional Theories of Physiological Measures Are Difficult to Test

In discussing the neural generators of the N170 component, Rossion and Jacques (2012) noted that different studies have yielded evidence for substantially different generator locations and that this may simply mean that multiple face-specific processes contribute to the scalp voltage during the time period of the N170. This same logic may apply to most scalp ERP components. In the case of the P3 component, for example, it would be surprising if only a single probability-sensitive functional process generated a voltage over the top half of the head between 250 and 600 ms. In other words, multiple processes likely contribute to the voltages that we record even when we use difference waves to isolate a specific effect, and this will make it very difficult to test hypotheses about which processes contribute to the recorded voltage. For example, if we know that context updating is eliminated under certain conditions, but we still find a P3 under these conditions because of other processes that also contribute to the P3, then we might incorrectly reject the hypothesis that context updating contributes to the P3.

We could partially solve this problem if we had more reliable methods for separating the mixture that we record at each scalp electrode from the underlying neural sources. However, this would not be a complete solution unless we assumed that a given region of cortex carries out only one function at a given point in time, which is certainly false. In primary visual cortex, for example, the initial feedforward wave of activity involves spatial filtering in simple cells, edge detection in complex cells, color processing in the cytochrome oxidase “blobs,” orientation contrast mechanisms in the “interblob” subregions, and so forth. Indeed, a single neuron may reflect the operation of multiple functional processes at a given moment in time. For example, the firing rate of a single neuron in area V4 may be simultaneously influenced by shape extraction processes, surround inhibition, and color constancy processes. In general, a key lesson of neuroscience over the past 50 years is that there is no 1:1 relationship between function and neuroanatomy at any scale.

Russ Poldrack, one of the deepest thinkers in neuroimaging, wrote an influential paper about fMRI research that addresses a similar issue, which he called the *problem of inverse inference* (Poldrack, 2006) (for an extended discussion of this problem in the context of ERPs, see Kappenman & Luck, 2012). In the context of neuroimaging, this problem is framed as follows: If brain activity has previously been observed in area X when process P is active, can we use the presence of activity in area X in a new experiment as evidence that process P was active in that experiment? This can be reframed in ERP terms: If component Y has previously been observed when process P is active, can we use the presence of component Y in a new experiment as evidence that process P was active in that experiment? The answer to both questions is no (unless we have additional evidence about brain area X or component Y).

The problem is that reverse inference is a case of the well-known logical error of *affirming the consequent*: if P entails X, X does not necessarily entail P. For example, if it is raining, then there must be clouds in the sky; however, the presence of clouds in the sky does not mean that it must be raining. Similarly, if previous evidence shows that the P3 wave occurs whenever the brain engages in context updating, a P3 could still occur under conditions when context updating is absent. Reverse inference is valid only when it is possible to say that X occurs if *and only if* P occurs (i.e., X never occurs without P). For example, this would be like saying that the P3 occurs if and only if the brain is engaged in context updating (which goes beyond saying that the P3 occurs if the brain is engaged in context updating). This is especially problematic for fMRI because the thousands of neurons in a given voxel will almost certainly be active for more than a single process. Thus, a BOLD signal may be seen in area X whenever process P is active, but it is unlikely that this BOLD signal will be seen *only* when process P is active.

The good news for ERP researchers is that because only a fraction of brain activity will yield a scalp ERP signal, we are more likely to be able to show that a given ERP component is present if and only if a given process is present (assuming that we can isolate that component). For example, multiple experiments show that the N2pc component is present if and only if attention is allocated to an object (Luck, 2012b). Consequently, when Michelle Ford and I found that an N2pc was present for conjunction targets and not for feature targets under certain conditions (Luck & Ford, 1998), we were reasonably justified in using reverse inference to draw the conclusion that there is at least one mechanism of attention that is necessary for conjunction targets but not for feature targets. However, you should note the phrase, "at least one mechanism of attention" in the preceding sentence. As discussed previously in the main body of chapter 3, we do not yet know the precise nature of the process that the N2pc reflects, and it might reflect multiple attention-related processes. Indeed, it may reflect an ancillary process that is triggered when attention is focused on an object rather than the attentional mechanism itself. Conclusions based on reverse inference will only be as precise as the link between the ERP component and the hypothetical process.

This brings up another key problem, which Emily Kappenman and I called the *problem of forward inference* (Kappenman & Luck, 2012). Specifically, it is harder than you might realize to test the hypothesis that component Y occurs if and only if process P is active. The difficulty arises because the whole reason we are seeking a physiological measure of process P is that we do not fully understand this process and therefore need a physiological measure to study the process. If we already understood process P, why would we need an ERP index of this process? However, if we don't already understand process P, we presumably don't have an ironclad way of knowing when process P is active. And if we don't know when process P is active, we cannot test the hypothesis that component Y occurs if and only if process P is active.

Getting Around the Problem of Forward Inference

There are three general ways of solving this problem. The first two involve lowering our aspirations and being satisfied with somewhat weaker or less precise conclusions that we might like. First, we might settle for showing that component Y *usually* occurs if and only if process P is *probably* active. Reverse inference could then be used in new experiments, but the conclusions of these experiments would be only probabilistic (e.g., "process P was probably active in condition A of this experiment"). This is not very satisfying, but progress in science often results from many probabilistic but converging results.

The second solution is to settle for conclusions that do not involve the specification of highly precise cognitive functions. For example, Geoff Woodman and I published a series of experiments showing that the N2pc component shifts rapidly between the left and right hemispheres when a visual search display contains potential target items on both the right and left sides (Woodman & Luck, 1999, 2003b). Because previous studies had shown that the N2pc is related to some kind of mechanism of attention, we were able to conclude that attention was shifting rapidly from item to item even though we could not say *exactly* what attentional mechanism was being indexed by the N2pc. Fortunately, any evidence that attention shifted rapidly from item to item in visual search was theoretically significant, even if we couldn't delineate the precise mechanism of attention we were measuring.

This second solution, like so many aspects of ERP research, depends critically on the ability to isolate a specific ERP component. The N2pc was isolated by looking at the contralateral-minus-ipsilateral difference. Because contralateral and ipsilateral were defined relative to the positions of potential target items, as defined by the attentional requirements of the task, we can be quite certain that the pattern of contralaterality we observed was associated with attention. That is, the difference wave was defined in terms of contralaterality with respect to to-be-attended items, so it would be difficult to come up with an alternative explanation that did not involve attention. In contrast, even if we could be certain that the anterior N2 component was perfectly related to conflict, finding that the voltage over anterior scalp sites between 200 and 300 ms is more negative in condition A than in condition B would not be sufficient to conclude with any confidence that the brain experienced greater conflict in condition A than in condition B (because this effect could be caused by some other component in the same time range).

The third solution to the forward inference problem involves directly confronting the problem and conducting a very involved program of research designed to assess the relationship between an ERP component and the functional process. This solution involves a "bootstrapping" approach (as in "pulling one's self up by one's bootstraps"). In this approach, you start by testing the most obvious predictions about the process of interest. That is, even if you do not have a complete theory of this process, you probably know enough about it to make some very simple predictions. For example, one of the first N2pc experiments I did in Steve Hillyard's lab involved varying the presence or absence of distractors. It seemed obvious that if the N2pc component reflects the filtering of distractors, then the N2pc should be eliminated if the target was not accompanied by any distractors. It also seemed obvious that the N2pc should be eliminated if the array contained multiple items but all of the items were identical. These predictions were confirmed (Luck & Hillyard, 1994b). Once your functional theory of a component has survived several of these obvious tests, it can be used provisionally as an index of the hypothesized process. If this leads to a coherent set of results about the hypothesized process, and these results are consistent with the results of experiments using other methods, then your confidence in the relationship between the component and the functional process will gradually increase. If you find discrepant results, however, then you will need to reappraise the link between the component and the process.

There are not many cases in which this approach has been used to solve the problem of forward inference. Perhaps the best example is the case of the ERN and its close cousin, the anterior N2. Researchers in this domain have gone back and forth between ERPs, fMRI, single-unit recordings, and behavioral experiments, leading to progress in understanding both the ERP components and the underlying cognitive and neural processes. But even if you are successful with this approach and can figure out the functional significance of a given ERP component, you will still be faced with the problem of knowing whether this component is responsible for the changes in voltage that you observe in new experiments. Thus, in many cases the best strategy is to design experiments that do not depend on determining the link between an ERP component and a specific functional process, as will be discussed in chapter 4.

I would like to stress that the problem of forward inference, and these three potential solutions to it, are not unique to ERPs. They apply to links between any type of physiological measure and functional processes. The major difference between ERPs and most other physiological measures is that ERP research must also confront the superposition problem. That is, it will be difficult for you to test predictions about the functional significance of an ERP component if you cannot isolate that component from the other components that are mixed together in the scalp recordings. Going back to the Donchin and Coles (1988) analogy, it would be hard for you to test hypotheses about the functional significance of the whirring noise if you can't separate the noise of the hard drive from the noise of the computer's fan. But the other problematic aspects of the forward and reverse inference also apply to fMRI, local field potential recordings, single-unit recordings, optical imaging, patch clamp recordings, and so forth.

Note

1. In their version of this analogy, Donchin and Coles (1988) described a floppy disk drive rather than a hard drive. I've updated the analogy because floppy drives have become increasingly rare, and some people might not appreciate the analogy. Of course, hard drives are beginning to be replaced by flash drives, so even this updated analogy will probably be outdated at some point.