Chapter 9

Experimental Design

All scientific research begins with a question. For cognitive neuroscience research, that question may be as broad as “In what brain regions is increased activation associated with successful retrieval from memory?” or as focused as “Does the functional connectivity between the frontopolar and anterior temporal cortices modulate the retrieval of an autobiographical memory?” From the experimental question, a researcher derives a research hypothesis, which is a statement about how a given manipulation should change some measurement. Hypotheses, like research questions, may be very general or very specific. An example of a general hypothesis in fMRI research would be the statement “Activation in the frontal cortex will be reduced in depressed individuals.” More specific hypotheses link fMRI data to other measurements; for example, “The activation in the middle frontal gyrus evoked by an n-back working memory task will decline proportionally to the degree of depression as measured by the Beck Depression Inventory.” The key characteristic of a hypothesis is that it is falsifiable, that is, an experiment could be conducted that could disprove it. More specific hypotheses can be tested more easily and are thus more informative.

To test a hypothesis, a scientist designs an experiment. Experiments, in the technical sense of the word, first manipulate some aspect of the world and then measure the outcome of that manipulation. The canonical example of an experiment is Galileo’s test of the effect of mass on acceleration due to gravity. Galileo speculated that an object’s acceleration due to gravity does not depend on its mass. To test this hypothesis, he dropped two balls of different mass from a great height and found that they fell at the same rate. In this experiment, Galileo manipulated the mass of the objects being dropped and measured the relative time needed for them to fall a given distance. In modern fMRI experiments, scientists often manipulate some aspect of a stimulus, such as whether a picture is of a face or an object, or whether a word is easy or difficult to remember, and then measure the change in BOLD signal within the brain. The way in which a scientist sets up the manipulations and measurements in an experiment is known as experimental design.

All well-designed experiments share several characteristics: they test specific hypotheses, they rule out alternative explanations for the data, and they functional connectivity A pattern of functional relationships among regions, inferred from common changes in activation over time, that may reflect direct or indirect links between those regions.

research hypothesis A proposition about the nature of the world that makes predictions about the results of an experiment. For a hypothesis to be well formed, there must be some experiment whose outcome could prove it to be false.

experiment The controlled test of a hypothesis. Experiments manipulate one or more independent variables, measure one or more dependent variables, and evaluate those measurements using tests of statistical significance.

experimental design The organization of an experiment to allow effective testing of the research hypothesis.
efficiency  The power of an experimental design to test a research hypothesis. Highly efficient designs can reject the null hypothesis even when the experimental manipulation has only a small effect.

variable  A measured or manipulated quantity that varies within an experiment.

independent variables (IVs)  Aspects of the experimental design that are intentionally manipulated by the experimenter and that are hypothesized to cause changes in the dependent variables.

conditions (or levels)  Different values of the independent variable(s).

dependent variables (DV$s)  Quantities that are measured by the experimenter in order to evaluate the effects of the independent variables.

minimize the cost of running the experiment. Advance planning to ensure good experimental design is especially important for fMRI experiments, given the significant resources they require in direct scanner costs and in the time spent by the experimenters, research assistants, and/or technologists in collecting and analyzing the data. If your experiment is inadequate for answering your hypotheses, all of that investment in time and money may be wasted. This chapter discusses the basic principles of experimental design for fMRI, returning at the end to a comprehensive discussion of design efficiency. Although designs differ in their advantages and disadvantages, there is one overriding principle: the best experimental design is the one that lets you investigate effectively your particular research question.

Basic Principles of Experimental Design

The fundamental element of an experiment is the variable, of which there are two types. Independent variables (IV$s) are aspects of the experimental design that are manipulated by the researcher. The choice of IV depends on the hypothesis to be tested. In Galileo’s experiment, the IV was mass, and there were two levels of mass (light and heavy). Different types of fMRI studies use different IV$s. In a study of visual perception, one could use the variable *stimulus category* by showing subjects different types of objects (e.g., faces, houses, tools). A study of attention might manipulate whether or not subjects pay attention to a given object, so that one condition could be *attended* and the other *unattended*. And in a study of long-term memory, one could train subjects using a list of words one week before putting them in the scanner, and then compare *remembered* words with *novel* words that were not previously learned. The different values of an IV are often called conditions or levels.

Dependent variables (DV$s) reflect the data measured by the experimenter. Different dependent variables provide different evidence for or against a hypothesis. Galileo compared how rapidly two objects fell; his dependent variable was relative time. Most fMRI studies use BOLD signal change as the primary dependent variable, although a few studies use other measures, as described in Chapter 12. It is important to recognize that a single hypothesis can be evaluated using multiple dependent variables. For example, a common experiment in physics teaching laboratories is to repeat Galileo’s experiment while taking photographs of the falling objects at regular time intervals using a strobe-light system, as shown in Figure 9.1. The photographs provide information about the dependent variable, distance, from which velocity can be calculated. Analogously, a single hypothesis about the brain could be tested using fMRI, event-related potentials (ERP), magnetoencephalography (MEG), or lesion studies. Different neuroscience techniques provide different dependent measures, which can together provide converging evidence for a hypothesis.

Figure 9.1  A simple experiment. By dropping two objects of different masses and photographing their positions at multiple time points, an experimenter can verify that the independent variable, mass, has no effect on the dependent variable, distance traveled over time.
**Thought Question**

The idea of converging evidence is a very important one in science. Why does the collection of data using different techniques (and thus different dependent variables) improve our ability to test research hypotheses?

Note that behavioral measures like response time and error rate can be considered as either DVs or IVs, depending on the context. In a psychological experiment, researchers collect behavioral data in order to examine the effects of the experimental manipulations, for example, when examining whether attention (IV) decreases response time (DV). However, if one is interested in the effects of errors (IV) on BOLD signal (DV), the behavioral data may become the manipulated factor. In the remainder of this text, we will define independent variables as those aspects of the experiment that serve as factors in an analysis, and dependent variables as those aspects that serve as data for the analysis.

Experimental variables, whether independent or dependent, may be categorical or continuous. A categorical variable can have one of a number of discrete values (Figure 9.2A). For example, if you want to map the hand regions of the motor cortex, you may set up an experiment in which the subject squeezes with either the left hand or the right hand. The IV in this experiment is hand, which obviously has only two values. But imagine that you are interested in measuring how activation in the motor cortex (DV) changes with the

![Image of categorical variables](image)

**Figure 9.2 Categorical versus continuous variables.** (A) A categorical variable is one that can take two or more discrete values. Thus, whether the subject squeezes her left hand or right hand is a categorical variable. (B) A continuous variable can take any of an infinite number of values, limited only by the precision of the measurement. How much force the subject exerts when squeezing her hand could be measured as a continuous variable. (C) Some continuous variables can be converted into categorical variables. Color, for example, is a continuous variable since it results from the frequency of visible light, which can take any of a range of values. However, we have names for discrete color categories in the continuous spectrum.
continuous variable A variable that can take any value within a range.
between-subjects A manipulation in which different conditions are assigned to different subject groups.
within-subjects A manipulation in which each subject participates in all experimental conditions.

pressure of squeezing (IV). Subjects could squeeze a sensor that measures how much force is exerted. In this experiment, force would be a continuous variable, because it could take any value within a range (Figure 9.2B). Categorical independent variables have been more commonly used in fMRI experiments, as they allow the use of treatment/control analyses, which are described in the next section. However, using continuous variables can in principle be much more powerful, and some experimental questions require them. Sometimes fMRI researchers discretize (i.e., convert from continuous to discrete) a continuous variable into a limited number of categories (Figure 9.2C), such as when classifying response times, usually measured in milliseconds, into the categories of “fast” or “slow.” Changing continuous measures into categorical variables simplifies many types of analyses, but at the cost of eliminating a potentially important source of variability in the data. As fMRI analyses have grown more sophisticated (see the following two chapters), the use of continuous variables has become increasingly prevalent.

An important distinction can also be made between two types of manipulations. In a between-subjects manipulation, different subject groups reflect different values of the IV. Note that the group difference may be some intrinsic qualifier, like males versus females or drug abusers versus abstainers, or it may be assigned by an experimental manipulation. In fMRI studies, between-subjects experiments are most common in examinations of the effects of a drug or disease state. To study the effects of schizophrenia on executive processing in the frontal lobe, a researcher could run two groups of subjects, one with the disorder and one without. Designs that have different groups of subjects do different tasks are used less commonly, because differences in the groups’ characteristics such as in age, gender, or education, could confound the results. For this reason, most fMRI studies use within-subjects manipulations, where each subject participates in all experimental conditions. As a rule of thumb, experimental designs should be within-subjects whenever possible.

These concepts apply not just to fMRI studies but to any research program. For additional discussion of the general principles of experimental design, we refer the interested student to the Chapter References, where several comprehensive texts are listed.

Setting Up a Good Research Hypothesis

Underlying any experimental design is a research hypothesis, which has the following basic structure: “Manipulating the independent variable (IV) will cause changes in the measurement (DV).” The hypothesis is validated if we manipulate the IV and the DV changes as expected, but it is falsified if the DV does not change. A hypothesis can be made more precise by specifying how IVs and DVs should relate to each other. For instance, “Increasing the IV should cause a decrease in the DV.” While hypotheses can be stated in many different ways, they all have this same underlying structure of cause and effect.

In fMRI studies, there are three distinct levels of research hypotheses, representing three different types of questions that can be asked (Figure 9.3). At the most specific level are hypotheses about hemodynamic activation in the brain. Such hypotheses reflect questions about the BOLD effect itself, without making inferences about its causes. Many of the studies of refractory effects discussed in Chapter 7 fall into this category, such as the 2002 study by Birn and colleagues that investigated changes in the linearity of the hemodynamic response in different brain regions. A second class of hypotheses addresses
Figure 9.3 Constructing research hypotheses. Hypotheses are statements about the relationships between independent and dependent variables. For fMRI experiments, there are three types of research hypotheses. The most basic are hemodynamic hypotheses, statements about hemodynamic activation measured by fMRI. More complex are neuronal hypotheses, which make claims about how underlying neuronal activity should affect fMRI data. Finally, psychological hypotheses attempt to relate some aspect of cognition to observed fMRI results. Psychological hypotheses are the most challenging to construct, but they can have the greatest influence on the study of the brain.

questions about neuronal activity. Since fMRI does not measure neuronal activity directly, researchers must estimate that activity by transforming the measured BOLD signal. An example of a neuronal hypothesis for fMRI would be: "Neuronal activity within the superior temporal sulcus increases when viewing a video of a walking person." Note that although the fMRI measurement is still the BOLD signal, the manipulation is restricted to a specific class of stimuli and the inference relates to the neuronal activity itself. The third type of fMRI hypothesis is the psychological hypothesis. We can use fMRI to answer questions about psychological processes like attention, memory, or perception. One important hypothesis that has been studied using fMRI is: "Encoding of items into memory and retrieval of items from memory are associated with activation in different hemispheres." Note that this hypothesis relies on very general concepts (encoding and retrieval) that are not uniquely defined. Researchers can (and do) disagree over what those terms mean.

Psychological hypotheses can be the most difficult to construct, but they are often the most influential. Consider a hypothesis about the organization of the visual system that was advanced by Ungerleider and Mishkin in 1982. They suggested that visual information might be processed in two distinct pathways, a ventral occipitotemporal pathway that processes object features ("what") and a dorsal occipitoparietal pathway that processes spatial properties ("where"). From this simple statement have come literally hundreds of neuroimaging and electrophysiological studies, along with extensions of the initial hypothesis to include dorsal and ventral divisions in the frontal lobe and even debates over exactly what sorts of spatial/object information are represented. This particular hypothesis has become sufficiently well supported that it forms the basis of a theory, a generalizable set of rules that shapes thinking on a topic, in this case the visual system. It is important to recognize that this influential idea began with a simple and falsifiable hypothesis. Psychological hypotheses are limited by how well we can define the concepts of inter-

theory An organized set of ideas that guides thinking on a topic and that can be used to generate a variety of experimental hypotheses.
A condition that contains the stimulus or task that is most relevant to the research hypothesis. Also called the task condition.

Control condition A condition that provides a standard to which the experimental condition(s) can be compared. Also called the baseline condition or nontask condition.

Epiphenomenal A secondary consequence of a causal chain of processes, but playing no causal role in the process of interest.

The “what/where” hypothesis depends on our intuitive understanding of the differences between spatial information and object information. Some researchers have suggested that spatial information includes how objects can be manipulated spatially, and thus the dorsal stream represents “how” information, not “where” information. The resulting debate has spawned new hypotheses about the organization of the visual system, changing the very terms used by the scientific community.

To test hypotheses, scientists set up experiments. Given the hypothesis “Manipulating the IV will cause changes in the DV,” the experiment must contain at least two values of the IV and must be able to measure changes in the DV. The simplest way to set up an experiment is to have two conditions that occur at different times. These are usually separated into an experimental condition and a control condition, which differ only in the effect of interest. The experimental condition is sometimes called the task condition, and the control condition is sometimes called the baseline condition or non-task condition. In Chapter 7 we discussed the first BOLD fMRI experiment, which was reported by Kwong and colleagues in 1992. They hypothesized that manipulating the amount of visual stimulation would change the BOLD activation level in the primary visual cortex. Their experimental condition consisted of bright flashing lights that the subjects watched using LED goggles, while their control condition was darkness. When different levels of BOLD activation in the visual cortex were measured between conditions, they attributed that difference to the IV of light stimulation.

Are fMRI data correlational?

A frequent criticism of fMRI data is that they are correlational, or epiphenomenal, implying that one cannot use them to make causal inferences and thus cannot conduct tests of experimental hypotheses (Figure 9.4A and B). This criticism is derived from the nature of the BOLD signal. As outlined in Chapter 6,
current theories of brain function assume that information processing results from neuronal activity. Of primary importance are axonal action potentials and dendritic field potentials, but other aspects, including synaptic changes and neurotransmission, are also critical. So, when a physiologist implants an electrode into the brain and measures changes in a neuron’s electrical potential, she assumes that such changes reflect some form of information processing, although the associated mental operations may be unknown. Hemodynamic changes, in contrast, do not necessarily reflect information processing. Remember from Chapter 7 that early studies showed that BOLD contrast could be evoked by physiological manipulations, such as CO₂ inspiration. These manipulations influenced the magnitude of the BOLD signal, presumably with little to no impact on computations in the brain.

What does it mean for the BOLD signal to be epiphenomenal, that is, merely correlated with neuronal activity? From a strong hypothesis-testing perspective, one could use fMRI data to test hemodynamic hypotheses, as described in the previous section, but not neuronal or psychological hypotheses. But nearly all fMRI studies investigate psychological questions! Fortunately for fMRI researchers, the correlational objection rests on an overly strict definition of hypothesis testing. All hypotheses are based on the principle that the experimental manipulation causes changes in the dependent variable. However, the chain of causation does not have to be fully elaborated. Consider a typical study evaluating the effect of a drug. To examine whether the drug has a beneficial effect, a researcher gives it to one group of subjects, while another group of subjects receives a placebo. If the experimental group does better than the control group on some measure, such as lower incidence of cancer or increased performance on a memory test, then the beneficial-effect hypothesis is supported. Such a result does not mean that the drug is a direct cause of the dependent measure; it could influence other factors (e.g., mood) that in turn cause the effect. Nor does it mean that no other manipulation could cause the effect. From this example, it is easy to recognize that all experiments, even perhaps those of low-level physics phenomena, have implicit causal structure.

In summary, correlational is not equivalent to meaningless. As critics correctly note, the mechanisms of BOLD activation are still not completely understood. But the inability to completely explain how neuronal activity leads to BOLD signal does not call into question that the two are related. Consider a simple analogy. You are standing next to a train track, waiting for a train. Within a few minutes, you hear the train’s whistle far in the distance. Not being an engineer, you do not know what causes the whistling sound. Nor do you know whether the whistle is needed for the train to move, or is completely unnecessary and epiphenomenal. Despite your profound lack of knowledge about the mechanism behind the whistle, you are certain about one fact: when you hear a whistle, it means that a train is coming.

Just as the whistle serves as a reliable predictor of the train, BOLD fMRI data serve as a reliable predictor of neuronal activity. We return to this issue in Chapter 14 by considering how to combine fMRI data with information derived from other techniques.

**Confounding factors**

If an experiment has only two conditions, experimental and control, it is critical that they be as similar as possible. If the conditions differ in only one property, then any change in the dependent variable can be confidently attributed to the change in that property. This process is known as subtraction, since one...
Figure 9.5 Selecting appropriate control stimuli. To study brain regions associated with face processing, you would want to manipulate the face-like characteristics of the stimuli in the experimental and control conditions. In the experimental condition, you could present a series of faces (A). However, many different control conditions are possible. (B) One option would be to present faces that have been transformed so that low-level visual properties are kept constant but the faces are no longer visible. Shown are images of faces that have been Fourier transformed, phase-scrambled, and inverse Fourier transformed. The same spatial frequency components are present, but they no longer form a face. (C) Another option is to present a series of simple objects in the control condition. The objects are visually interesting, like the faces, and have smaller parts that are nameable.

can subtract the value of the dependent variable in the control condition from its value in the experimental condition to quantify the effect of the manipulation. But if the conditions differ in more than one way, then there could be multiple explanations for experimental effects. Any factor that varies with the IV in an experiment is known as a confounding factor. Perhaps the most important aspect of experimental design, but the most difficult to master, is selecting good experimental and control conditions in order to minimize confounding factors.

FMRI studies with psychological hypotheses are particularly susceptible to confounding factors, since the concepts they address are often difficult to define. To understand why, consider the hypothesis that face perception relies on the fusiform gyrus within the inferior temporal lobe. The experimental condition seems obvious: present photographs of human faces (Figure 9.5A). But what is the appropriate control condition? One option would be to simply show nothing, making the design analogous to the Kwong study described above. The experimental and control conditions would thus differ in the intended IV, in that the experimental condition would present faces while the control condition would not, but they would also differ in other factors. In this case, confounding factors would include brightness, presence of edges, nameability, and visual interest, among many. Another possible control condition would be to present faces that have been transformed in some way, so that parameters like brightness and spatial frequency composition are similar between the two conditions (Figure 9.5B). Yet face perception means more than just processing of the physical properties that make up the face. It also refers to seeing an image as a face as opposed to some other type of object. This psy-
chological interpretation suggests another possible control condition, the presentation of random objects of generally similar complexity to the faces (Figure 9.5C). Such a comparison would identify areas of the brain that respond more to the faces than to the other similar objects.

An even more insidious type of confounding factor in fMRI studies is the hidden causal factor. According to the oft-repeated mantra, “Correlation does not imply causation.” Just because event A occurs at the same time as event B does not mean that A causes B or that B causes A. Failure to heed this warning can lead to the confirmation of bad hypotheses. A classic example can be found in the link between ice cream consumption and violent crime: both are highest in summer months and lowest in winter months. Does this mean that eating ice cream causes crime, or vice versa? Of course not. Both variables are influenced by a third factor: temperature. This example will seem remedial to many students, but its basic logic holds in more complex fMRI studies. Suppose that you are studying the effects of alcohol on motor cortex function. Your subjects watch a computer screen on which the letters “L” and “R” flash in a random sequence, and they squeeze their left or right hands in response to the corresponding letters. You find that the BOLD signal in the motor cortex is reduced in subjects who drank alcohol compared with those who drank water. What do you conclude? One possibility, that alcohol reduces neuronal activity in the motor cortex, seems reasonable from the data. However, other interpretations should be considered. For example, the subjects may make many more mistakes under the influence of alcohol and squeeze their hands at the wrong times or not at all. The reduced BOLD activation may thus result from poor behavioral performance, not directly from the alcohol. Note that you cannot identify which factor, alcohol or behavioral performance, causes the reduced response with this single experiment. Additional experiments would be necessary to determine the true cause of the effect.

Confounding factors can be minimized if they are not correlated, within your experiment, with the independent variables of interest. The most basic approach to ensuring that this is the case is simple randomization. For example, making your experimental trials occur in a random order can eliminate many sorts of confounding factors related to response preparation, stimulus expectation, and task-switching. Note that, as is discussed in Box 9.2 at the end of this chapter, recent research has identified methods of randomization that take into account properties of the BOLD hemodynamic response to identify optimal sequences for trial presentation. When factors cannot be made completely random, or when randomizing might actually introduce a confounding factor (e.g., because of small numbers of subjects or events), scientists often try to ensure that a potential confounding factor is equally present for all conditions. This is called counterbalancing. If your experiment is comprised of three runs with different task instructions, the effects of practice or fatigue could be important. For some tasks, subjects may get better over time (i.e., practice effects), and thus their performance may be better in later runs. For other tasks, subjects might tire over time (i.e., fatigue effects), and their performance may worsen as the experiment goes on. To ameliorate these problems, you could counterbalance the order of runs across your subject groups: some are given the tasks in the order 1-2-3, others in the order 2-3-1, and the remainder in the order 3-1-2. By introducing randomization or counterbalancing into your design, you can increase the chances that confounding factors influence all conditions similarly.

A good way to identify confounding factors is to participate in your own experiments as a pilot subject. You may recognize an unexpected confound-
ing factor when you adopt a different strategy than expected, or when you find that the task is too easy or too difficult. Although one cannot always predict all the possible confounding factors, the costs of fMRI experiments in time and money provide ample incentive for good experimental design. The best designs enable the researcher to efficiently answer the questions of interest while requiring a minimum number of experimental subjects and experimental trials per subject.

All of the concepts discussed so far in this chapter are applicable to all areas of science, not just to fMRI experiments. Regardless of the topic, method, or discipline, researchers should always question their choices of independent variables, dependent variables, and hypotheses (Figure 9.6). Thoughtful experimental design is the cornerstone of the scientific method.

Good Practices in fMRI Experimental Design

Before discussing specific types of fMRI experimental design, we want to introduce some guidelines that are relevant to all fMRI studies. As emphasized throughout this textbook, it is very easy to conduct an uninteresting, underpowered, or uninterpretable fMRI study—and the most common cause of problems is poor experimental design. Conversely, attention to experimental design at the outset of a research project will maximize the chances of success. We advocate six basic (and non-exhaustive) rules for fMRI experimental design:

1. **Evoke the cognitive (or motor, perceptual, mnemonic, etc.) processes of interest.** This first rule may seem trivial, but it reflects a very real problem in fMRI research. It is very simple, especially for novice investigators, to create a seemingly well-designed fMRI task that fails to generate the very cognitive processes in which they are most interested. When designing an experiment, the first question you should ask is “What will my subjects be doing during this task?”

2. **Collect as much data as possible from each subject.** Ensure that, when your subjects are in the scanner, time is spent on the key task conditions, not wasted on nonessential aspects of the experiment. Without knowing the size of the effect that your manipulation will evoke, it is difficult to answer the question “How many trials do I need?” That said, collecting more data always gives you a better chance of obtaining interpretable results.

3. Moreover, if your design involves the integration of data across subjects, as is the case for most fMRI research, then **collect data from as many subjects as possible.** This is especially critical for comparisons between groups of subjects (e.g., younger vs. older adults), for which both intra- and inter-subject variability can compromise results.

4. **Choose your stimulus conditions and the timing of their presentation to evoke maximal changes in the cognitive processes of interest, over time.** Usually, this involves clumping stimuli so that there are some periods with continual engagement of the processes of interest, and other periods with no engagement of the processes of interest. Doing this increases the efficiency of experiments involving the detection of activated regions.

5. **Organize the timing of experimental stimuli so that successively elicited processes of interest are minimally correlated with each other,**
over time. This often requires using a variable interval between successive events, especially when those events occur relatively rapidly or as part of a complex task. Doing this increases the efficiency of experiments involving estimations of the time course of activation.

6. Where possible, obtain measurements of your subjects' behavior that can be related to the fMRI activation. This allows researchers to test more specific hypotheses about brain function, compared with simple comparisons of experimental conditions. Depending on the research question, these measurements could be obtained during the fMRI session itself (e.g., task performance), they could be part of a parallel behavioral experiment (e.g., subsequent memory effects), or they could be collected as independent assessments of some trait (e.g., intersubject variability).

Good experimental design is probably the most critical aspect of fMRI research. In the following sections we introduce some broad approaches to designing fMRI studies. But the topic of experimental design extends far beyond the scope of this single chapter. For a more extensive consideration of this subject, we refer the reader to the excellent articles by Henson, Liu, Poldrack, and Wager (cited at the end of the chapter).

**Blocked Designs**

The simplest way to evaluate the effect of the independent variable on the dependent variable is to compare an experimental condition (in which the IV is present) to a control condition (in which it is absent or at a lower level), using the logic of subtraction described in the previous section. For example, imagine that you want to investigate whether listening to music improves studying for examinations. Your subjects listen to a list of 20 words read one at a time. For some subjects, during the first ten words there is music playing in the background, while during the last ten words the room is quiet. To counterbalance the order of presentation, other subjects listen to music during the last ten words but not the first ten. In this experiment, the trials from each condition are grouped together in time to form blocks, as shown in Figure 9.7A. As a general definition, the independent variable is kept at a constant level throughout a block, and transitions between blocks represent changes in the level of the independent variable. Here there are two blocks, music and quiet, combined within a single blocked design. The basic analysis of any blocked-design experiment, whether fMRI or not, involves comparing the dependent measure in each block condition. For example, your subjects might remember, on average, eight of the words that they heard while music was playing but only six of the words that they heard when it was quiet. Unsurprisingly, such blocked designs, which are simple to create and straightforward to analyze, dominated the early years of fMRI (see Chapter 7).

To understand why blocked designs were first used, one must consider the context in which these early experiments took place. In the early 1990s, the magnitude of the BOLD change caused by neuronal activity was still unknown, and thus researchers adopted long block intervals to ensure that sufficient neuronal activity would be generated to evoke a measurable BOLD response. In addition, long task blocks had been necessary for PET imaging, which measures the total number of emission events following injection of a radioactive tracer (see Box 7.1). In a typical PET experiment using O¹⁵ to measure blood block A time interval that contains trials from one condition.

blocked design The separation of experimental conditions into distinct blocks, so that each condition is presented for an extended period of time.
Figure 9.7 Basic principles of blocked designs. In a blocked fMRI design, the experimental tasks are separated into long-interval blocks. (A) A simple blocked design, in which subjects read a list of words presented one at a time. During the first block of ten stimuli, the subjects hear music playing, while during the second block of ten stimuli, no music is heard. Note that although each of these blocks contains multiple individual stimuli, in most blocked-design analyses it is assumed that the cognitive processes of interest are constant throughout the block. The most common blocked design alternates between two conditions (B), allowing identification of the difference in fMRI activation between them. For some research questions, a rest or baseline condition is introduced between the two blocks (C) so that activation that is independent of both conditions can be measured.

Flow, a tracer would be injected and then the subject would perform one condition of the task for 60 to 90 s. Then there would be a second injection of the tracer, followed by 60 to 90 s of a second condition. A similar blocked approach was naturally adopted for the first fMRI studies, based on the idea of comparing steady-state activation in one task to steady-state activation in another. Even now, when much more is known about the construction and analysis of complex task designs, the simple blocked design remains an important part of fMRI.

**Setting up a blocked design**

The first issue to consider when creating a blocked design is the research question itself. Some experiments require long task blocks because the process of interest cannot be modulated over short intervals. If one is interested in studying vigilance or sustained attention, one could compare 30 s blocks in which subjects are concentrating on a task with 30 s blocks in which the subjects are not concentrating. Since active concentration may take some time to engage and disengage, using a blocked design will improve the subjects' ability to per-
form the task. Conversely, some experiments cannot use blocked designs due to the transience of the neuronal activity. Detection of infrequent targets, as in the common "oddball" or "n-back" paradigms, provides a good example. Imagine that you are watching a series of letters flashing rapidly on a computer screen. Your task is to press a button whenever you see an "X," which only appears five percent of the time. The oddball "X" cannot be presented repeatedly within a block, as that would change how subjects process it. An event-related or mixed design would instead be necessary.

Assuming that a blocked design is practical for an experiment, the researcher must next choose the experimental conditions and determine the timing of the blocks. The former requirement relates to the IV, in that conditions must be selected that maximally influence the desired IV without introducing confounding factors. The latter requirement relates to the DV, since the properties of the hemodynamic response determine the length of the blocks and whether there should be spacing between them. The choice of conditions for the different blocks relates in an important way to the goals of the experiment. Imagine that you are interested in whether or not nouns and verbs are processed in different areas of the brain. One obvious design would involve two conditions, nouns and verbs, each consisting of a series of words presented one at a time. Each condition could be presented for 30 s, and the conditions could alternate for the duration of the experiment. This alternating design (Figure 9.7B) is optimal for determining which voxels show differential activation as a function of the independent variable (i.e., the difference between the conditions). However, it does not provide any information about voxels that are active in both conditions or about the response to a single condition in isolation. To gain information about the independent responses to each condition requires additional control blocks (Figure 9.7C). Control blocks in which the subject does nothing, such as watching a blank screen where there is nothing to read, are also called null-task blocks. (Note that even null-task blocks may engage robust cognitive processing, as considered in Box 9.1.) Additional conditions require additional time, however, and should not be added unnecessarily. Therefore, whenever you choose which conditions to include in a blocked design, you should begin by evaluating whether a simple alternating design would be sufficient for answering your research questions.

After deciding on your experimental conditions, you should next consider the timing of your task blocks. Functional MRI experiments generally involve blocks ranging from about 10 s to about 1 minute. (Note that the use of very short blocks, especially when presented in randomized order, effectively changes the design to event-related.) Within that large range, the experimenter has considerable flexibility in the choice of timing parameters. Most important to consider is the effect of block length on the experimental task. Are there time constraints that preclude very short or very long blocks? In many working memory experiments, for example, subjects must rehearse a changing set of items over time. If the block is too short or too long, such a task may be too easy or too difficult. As for many psychology experiments, fatigue effects (and to a lesser extent, practice effects) should be considered. Very demanding tasks may be difficult to sustain over extended periods of time, and subjects may do worse at the end of a long block than at the beginning. In general, block length should be chosen so that the same mental processes are evoked throughout.

For most purposes, block length should be kept constant for all of the conditions. Remember that in an alternating design, the primary statistical analy-
BOX 9.1 Baseline Activation in fMRI

The basic assumption of blocked designs is that block-related changes in BOLD signal result from differences between the experimental conditions. In the usual subtractive approach, there are two conditions: task and control. The task condition is assumed to consist of all of the neural processes present in the control condition, along with additional processes of interest. Consider the results from the following early fMRI experiment, as reported by Binder and colleagues in 1999. During task blocks, subjects listened to sequences consisting of low and high tones and pressed a button if a given sequence included two high tones (e.g., L-L-H-L-H). In control blocks, the subjects lay still in the scanner with their eyes closed. Each block lasted 24 s. Not surprisingly, the tone task evoked more fMRI BOLD activation than control blocks within the auditory, prefrontal, parietal, and motor cortices, among many other regions. These areas reflect regions that are associated with perception, decision, and response aspects of the task. The authors then looked for brain regions that were less active during the task blocks than during the rest blocks. Surprisingly, a set of specific regions (now called the default network) showed increased activation during rest, a finding since confirmed by many studies using both fMRI and PET. These results collectively suggest that the assumptions of the subtractive method may be flawed. Some aspects of cognition may actually be inhibited during performance of psychological experiments, such that cognitive processes present in the control condition may not be present during the task condition. In the last few years, interpreting deactivations, or decreases in hemodynamic activation, has become an area of considerable interest.

What does it mean for fMRI activation to decrease during an experimental task? In answering this question, it is critical to recognize that blocked designs only provide information about the relative difference between two conditions, not about absolute levels of activation. Without a clear baseline to which both conditions can be compared, several different types of changes in absolute activation could result in similar changes in relative activation (Figure 1). Gusnard and Raichle, in a comprehensive review of many research studies published in 2001, suggest that the appropriate baseline condition for functional neuroimaging should be defined using the oxygen extraction fraction (OEF), which is largely stable across the brain (Figure 2). Even though some areas of the brain have greater blood flow than others, and some areas have higher oxygen requirements than others, the proportion of oxygen that is extracted when the subjects are resting with their eyes closed is spatially uniform, with only a few exceptions. Remember from Chapter 7 that the OEF decreases as part of the BOLD response.

default network A set of brain regions whose activation tends to decrease during the performance of active, engaging tasks, but to increase during conditions of resting and reflection.

deactivations Decreases in BOLD activation during task blocks compared with nontask blocks.

oxygen extraction fraction (OEF) The proportion of available oxygen that is removed from the blood.

---

Figure 1 Possible origins of increases and decreases observed in fMRI. When experimental and control tasks are compared using a blocked design, there are several possible causes of observed increases or decreases in hemodynamic activation. First, an increase in activation during the experimental task could be observed when both tasks are either above baseline (A) or below baseline (B). Likewise, decreases in activation during the experimental task could be observed when both are above baseline (C) or below baseline (D). Note that in (A) and (C), both tasks have a positive effect compared with baseline, while in (B) and (D), both tasks have a negative effect compared with baseline. If one task is above baseline and the other below (E), comparisons of the tasks with each other would yield a large effect, but the effect in relation to baseline would be indeterminate. (After Gusnard and Raichle, 2001.)
BOX 9.1 (continued)

Figure 2. The oxygen extraction fraction (OEF) as a possible baseline for brain activity. Gusnard and Raichle suggest that the OEF, the proportion of available oxygen that is extracted from the blood, is highly stable across the brain and represents a good baseline for brain activity. Shown are four axial slices reflecting data from an experiment with 19 adult subjects. Also shown here are relative cerebral blood flow (CBF) and cerebral metabolic rate for oxygen (CMRO₂): the ratio between these quantities gives the OEF. The arrows indicate the only regions of increased oxygen extraction relative to the remainder of the brain. These are in the visual cortex and likely reflect the fact that the subjects in this data set had their eyes closed. The baseline for these regions may reflect open eyes and normal visual stimulation. (From Gusnard and Raichle, 2001.)

due to an overcompensatory increase in blood flow. Decreases from the baseline OEF indicate increased neuronal activity, whereas increases indicate decreased neuronal activity. Thus, a deactivation observed in an fMRI experiment might reflect a counterintuitive increase in metabolic activation within that region during a resting or inactive state, compared with during the performance of some active task. (However, see the recent article by Morcom and Fletcher for some important caveats for this and related conclusions.)

Across a wide range of fMRI studies, there has been remarkable consistency in the regions that evince deactivations during experimental tasks (or activations during non-task periods). These regions include the medial prefrontal cortex along the medial frontal gyrus, the posterior cingulate or precuneus, the lateral parietal cortex along the angular gyrus, and parts of the temporal lobe. Moreover, these regions exhibit a high degree of functional connectivity, such that moment-to-moment fluctuations in the activation of one region predict similar fluctuations in the other regions. Greicius and colleagues, in a 2003 article, collected fMRI data while subjects were lying in the scanner but performing no experimental task (i.e., a null-task session). They found that changes in the activation of the posterior cingulate are mirrored by similar changes in the medial prefrontal cortex and in the lateral parietal cortex, even in the absence of any external stimulus. Moreover, the activation of the posterior cingulate cortex was inversely correlated with activations in regions of the lateral prefrontal cortex associated with the goal-directed control of behavior. More recently, researchers have tracked changes in the functional connectivity among these regions between different stages of human development. A 2008 study by Fair and colleagues demonstrated that young children have reduced functional connectivity, particularly between the medial prefrontal and medial parietal cortex, compared with adult subjects. Functional connectivity between these same regions also declines in older adults, as shown in an elegant 2007 study by Andrews-Hanna and colleagues. Notably, this decline was greatest in individuals who also showed the highest levels of degeneration in connecting white-matter tracts, as measured using diffusion tensor imaging (see Box 5.1). This suggested that intact fiber connections between these regions are necessary for their functional connectivity.

What functions might this seemingly arbitrary set of regions serve? One possibility is that they may play important roles in the monitoring of external stimuli. Both the medial and lateral parietal cortices have been implicated in spatial and attentional processes. However, neuroimaging and single-unit studies indicate that these regions are not associated with attention to expected stimuli. Instead, they seem to be more associated with peripheral or unexpected events, consistent with the idea that they are part of a generalized monitoring system. The ventromedial frontal cortex, in contrast, is typically associated with emotional processing, including assessments of the likely reward consequences of future actions. Together, these sorts of monitoring processes can be considered examples of "self- (continued on next page)
BOX 9.1 (continued)

directed”, “stimulus-independent”, or “introspective” thought (see the References for several articles discussing these ideas).

To experience self-directed thought, close your eyes and relax for at least 10 s. If you are like most people, you will feel an initial sense of withdrawal from the world around you, followed by a growing sensitivity to external stimuli. You will notice sounds that had previously been outside your awareness. You will become sensitive to heretofore unnoticed muscle tension or joint pain. In short, the baseline state of brain activation is very different from that evoked by a demanding task, like reading this book. This difference was recognized by William James, who described introspection as follows: “When I try to remember or reflect, the movements [in the mind] in question, instead of being directed toward the periphery, seem to come from the periphery inwards and feel like a sort of withdrawal from the outside world.” (1890, p. 300)

Here, James emphasizes the essential difference between active and reflective states. The former is goal-directed, aimed at changing the surrounding environment or one’s place in it. The latter is self-directed and passive, seeking information about the environment. James also notes his awareness of particular body motions during introspection, and how those physical movements might relate to the movement of thoughts. As in much of James’s writing, his well-reasoned reflections anticipate the present discussion of baseline brain activation by more than a century. Rest or baseline conditions are not absent of mental processes. Instead, they contain particular types of processes associated with reflection, daydreaming, self-assessment, bodily attention, and emotion. Recent work by Mason and colleagues, for example, links stimulus-independent thought (e.g., mind-wandering) to activation of default network regions (Figure 3). When designing an fMRI experiment, you must account for

Figure 3 One of the potential interpretations of increased activation during non-task-related conditions is that it reflects “stimulus-independent thought,” such as daydreaming and mind-wandering. To test this interpretation, Mason and colleagues trained subjects on working-memory tasks. On blocks of stimuli that were well-practiced, subjects became more likely to think non-task-related thoughts, compared to novel blocks of stimuli that required focused attention. The researchers compared fMRI activation during practiced and novel blocks. Shown circled at left are regions with greater activation during practiced blocks; these included the medial frontal cortex (A) and the posterior cingulate cortex (B). As shown on the scatterplots to the right, those subjects who were more likely to engage in mind-wandering (x-axes; z-scores on an independent daydreaming scale) exhibited the greatest activation in each of these regions (y-axes). The researchers conclude that these regions play an important role in stimulus-independent thought. (From Mason et al., 2007.)
reflective processes in your choice of experimental and control conditions. If the control condition has no explicit task requirements, subjects will naturally begin thinking about how they are doing in the experiment, what they will have for dinner, which friends they will see this evening, or even a dull pain in their lower back that they only now are beginning to notice. For this reason, rest or null-task conditions are not recommended, as the sole control condition. Instead, conditions should be chosen so that subjects are always performing some active task or attending to a changing environment. For example, if the experimental condition consists of judging the familiarity of remembered words, consider a control condition where subjects read words and indicate whether they are presented in capital or lowercase letters. Both conditions require attention and decision processes, thus precluding activation in the baseline system, but only the former invokes memory processes. Another suggested procedure is to explicitly include a null-task condition within your design to provide a baseline. By doing so, you can evaluate whether relative differences between your conditions reflect differential increases above baseline or an effect for one condition but not the other (see Figure 1). Understanding the effect of baseline processing on fMRI data is important for any researcher, but it is especially critical for those who use blocked designs.

As a final point of consideration, several theorists have argued that the pervasive co-activation of default-network regions may reflect something much more fundamental than conscious but undirected thought. They point to phenomena like the presence of functional connectivity between these regions in patients under general anesthesia, along with the sorts of broad lifespan changes in that connectivity described earlier in the box. The sorts of fundamental processes that these regions might support remains mysterious, but there have been some speculations. These regions may form predictions about future mental operations, coming online when immediate task demands cease. They may increase the sensitivity of other brain regions to external stimuli, effectively turning up the gain on stimulus-response links elsewhere in the brain. Or, they may act to consolidate representations of past events, whether at the level of overt memories or learned behaviors. Whether one of these or another property accurately describes the collective function of these regions will remain an important topic for future cognitive neuroscience research.

sis evaluates differences between the two conditions. Even if one condition is labeled as the task and the other is called baseline, they are equally important for the statistical comparison. Statistical comparisons between two conditions are determined by the magnitude of the difference between the conditions compared with the variability within conditions, as seen in Figure 9.8. Since the standard deviation of an fMRI time course decreases with the square root of the number of time points (e.g., in a block), the overall standard deviation will be largest when one block is very long and the other very short (i.e., when the former has low variability, but the latter very high variability), while it will be smallest when the blocks are of equal length (e.g., both with intermediate variability). So, for optimal statistical power, the blocks in an alternating design should generally be of equal length. However, if more than two conditions are used, then unequal block lengths or block numbers may be beneficial. If a primary comparison is something like the combination of condition 1 and condition 2 versus condition 3, then it may be worth making condition three twice as long as the others. This often occurs in designs that use a null-task block along with two experimental conditions (i.e., 1-3-2-3-1-3-2-3, etc). Also, if additional analyses will examine responses to individual events within an experimental block, as with the mixed designs described later in this chapter, then that block may be lengthened relative to a control block.
Figure 9.8 Within-conditions and between-conditions variability in blocked fMRI data. The goal of experimental design is to maximize the variability in the data that is due to the experimental manipulation (i.e., the between-conditions variability) while minimizing other sources of data variability (i.e., within-conditions variability). If the former is large compared with the latter, effects of interest can be identified.

**Advantages and disadvantages of blocked designs**

Though simple, blocked designs can be extremely powerful. For evaluating the strengths and weaknesses of an experimental design in fMRI, we consider two factors: detection, or knowing which voxels are active, and estimation, or knowing the time course of an active voxel. These factors correspond roughly to spatial and temporal resolution. Detection power depends on the total variance in BOLD signal introduced by the experimental design, while estimation efficiency depends on the randomness of stimulus presentation. A central principle of fMRI experiments is that a design that is good at detection may not be good at estimation, or vice versa. (We return to these issues in the extensive Box 9.2 later in this chapter.)

Blocked designs are very good for detecting significant fMRI activation. The detection power of a blocked design is determined by the balance between two factors. First, the difference in BOLD signal between conditions should be as large as possible. Figure 9.9 shows a simulation of how the measured BOLD activation changes as the length of the blocks changes from very long (40 s) to very short (2 s). With long block lengths, a very large response is evoked during the task blocks and the response returns to baseline during the nontask blocks. Thus, there is maximal variability between the blocks. If the block length is sufficiently short (i.e., less than about 10 s) then the hemodynamic response cannot return to baseline during the nontask blocks, and the BOLD amplitude will be reduced. This reduces the total variability in the data, which in turn reduces the experimental power. Reducing the block lengths to extremely short durations, such as only a few seconds, would lead to almost no difference in BOLD signal between task and nontask conditions. In summary, the use of long block intervals provides for maximal BOLD amplitude changes between conditions.

Second, the signal to noise ratio should be maximized at the task frequency. (Note that the task frequency is simply the inverse of the total task period; for example, if your experiment alternated a 20 s task block with a 20 s rest block,
Figure 9.9 Effects of block interval on the fMRI hemodynamic response (BOLD signal). These charts show simulated fMRI hemodynamic responses of voxels active only during the task block of an alternating on/off design. The duration of each block is shown in the upper right corner of each graph. Note that as the block duration shortens below the length of the fMRI hemodynamic response (about 10 s), the response does not return to baseline. At very short block durations, there will be little or no difference between fMRI signal during active and inactive blocks. Note that the scales of the y-axes are reduced for block lengths of six seconds or less.

your task frequency would be $1/40$ Hz.) The noise in a BOLD time course has its highest power at low frequencies and lowest power at high frequencies. For example, at very low frequencies, there can be significant scanner drift due to problems with the scanner hardware. If your design has very long (e.g., 180 s) blocks, it will be difficult to know whether signal changes from one block to the next result from the experimental manipulation or from low-frequency noise. As the block length is reduced, the task frequency increases and thus the design is more immune to low-frequency noise.

Together, these factors indicate that the noise at the task frequency will be greatest with relatively long block lengths and smallest with relatively short block lengths. As a rough guideline, block lengths of approximately the duration of the hemodynamic response (i.e., 10 s to 15 s) provide large signal changes while reducing noise at the task frequency to an acceptable level. However, depending on the spectrum of the noise, detection power may increase at even shorter block intervals of 6 s to 8 s (see the 1996 article by McCarthy and colleagues for an example). Longer blocks are often required for experiments that test cognitive processes like memory and attention, since it is difficult to ensure that those processes begin promptly. If design constraints necessitate the use of very short block periods, those blocks should be treated like single events and their order should be randomized. This procedure is described in the section on event-related designs.

While their detection power can be very good, blocked designs are relatively insensitive to the shape of the hemodynamic response. We can understand scanner drift Slow changes in voxel intensity over time.
superposition A principle of linear systems that states that the total response to a set of inputs is equivalent to the summation of the independent responses to the inputs.

Understand this insensitivity by returning to the idea of superposition, which was introduced in Chapter 7. Setting aside refractory effects for the moment, the hemodynamic response to two identical stimuli presented in succession is equal to the sum of the individual responses. As more and more stimuli are presented in succession, each contributes to the total hemodynamic response. With block lengths of about 10 s or more (i.e., longer than the duration of the hemodynamic response to a single stimulus), every time point within the block contains contributions from multiple stimuli, with each contribution at a different phase. The combined hemodynamic response thus rises rapidly at the onset of the task, then remains at a plateau until the end of the block.

Since the plateau value represents contributions from all phases of the hemodynamic response, the particular shape of the response does not matter. The yellow lines in each part of Figure 9.10 show four hypothetical hemodynamic responses, each with a different shape. The other colors show how the BOLD signal would change as more events are presented successively within a block. All four hypothetical responses have the same total signal amplitude, as measured by the areas under the curves. Consider the standard hemodynamic response shown in Figure 9.10A. As the length of the response block is

Figure 9.10 Insensitivity of blocked designs to the shape of the hemodynamic response. Each set of curves shows the simulated fMRI signal measured from blocks of 1, 2, 4, 8, 16, or 32 one-second stimuli. In (A), the base hemodynamic response shown in yellow has a standard form. As the number of stimuli in the blocks increases to 16 or more, the hemodynamic response reaches a plateau. Now consider the triangular waveform shown in (B). It is narrower than that in (A), so the response is different for small numbers of stimuli. But as more stimuli are averaged, the response approaches that of (A). Similar results can be seen for (C) and (D), which have very differently shaped hemodynamic responses. In fact, the hemodynamic response in (D) consists of the numerical values of the response in (A), but scrambled in a random order. So, if a single stimulus were present in the block, the BOLD data would look nothing like the standard hemodynamic response. Yet as increasing numbers of stimuli are averaged, the combined response will approach that of (A).
increased from two to 32 events, with each event being one second in duration and separated by two seconds, there is a consistent and smooth increase in overall signal amplitude, reaching a plateau at block lengths of about 12 s or longer. Now suppose that the hemodynamic response has a simple triangular form (Figure 9.10B). Obviously, this form differs considerably from that of the canonical hemodynamic response. But as the length of the block increases, the total hemodynamic response in panel B becomes more and more similar to that in panel A. As can be seen in the subsequent panels, the insensitivity of blocked designs to hemodynamic shape would hold if the hemodynamic response had two peaks (Figure 9.10C) and even if its values were completely randomized over the response duration (Figure 9.10D). For the same reasons, designs that use long stimulus blocks are also relatively insensitive to changes in the timing of the hemodynamic response; they would only identify effects associated with the onset of the first stimulus in the block.

Insensitivity to the shape and timing of the hemodynamic response has both advantages and disadvantages. The primary advantage is that it makes experimental analyses extraordinarily simple. When blocked, any hemodynamic response can be robustly modeled using a smoothed trapezoidal shape consisting of a rise, a plateau, and a fall. To evaluate the effect of the IV, the magnitude of the BOLD response during the task period can be compared with the response during a baseline period. If there is no effect, the random differences in activation between the blocks would follow the t-distribution, so the magnitude of the t-statistic reveals the significance of the effect. Other analyses are also possible, including power spectrum analyses, correlations, and explicit modeling of block waveforms (see Chapter 10). Balancing this advantage is a loss of estimation power. Imagine that you ran an experiment and recorded the data from the 32-stimuli-curve in Figure 9.10A. You would be unable to estimate whether the hemodynamic response looked like those in panels A, B, or C, because they all resulted in nearly identical data. In fact, with sufficient noise, even the random hemodynamic response in panel D would be impossible to distinguish from the others.

**Thought Question**

Imagine that you recreated Figure 9.10 using a sample hemodynamic response that is delayed by a few seconds. How would the amplitude and latency of the BOLD signal measured from that block be changed?

In summary, blocked designs are simple and powerful. They are easy to create and can be easily explained to others. If the experimental and task conditions are chosen carefully, then the analysis is very straightforward. Blocked designs are good at detecting voxels with significant activation, and they can identify a wide-range of task-related changes, regardless of any variation in the timing and shape of the BOLD signal. However, because the experimental condition is extended in time, it may evoke highly heterogeneous neuronal activity, making some tasks inappropriate for blocked designs. Blocked designs are also not useful for estimating the time course of activation in active voxels.

**Event-Related Designs**

The second major type of experimental design in fMRI is the event-related design. The central assumption of an event-related design is that the neural event-related design The presentation of discrete, short-duration events whose timing and order may be randomized.
event A single instance of the experimental manipulation. Also known as a trial.

interstimulus interval (ISI) The separation in time between successive stimuli. Usually refers to the time between the end of one stimulus and the onset of the next, with the term “stimulus-onset asynchrony” (SOA) used to define the time between successive onsets.

electroencephalogram (EEG) The measurement of the electrical potential of the brain, usually through electrodes placed on the surface of the scalp.

time-locking Synchronization of analyses to events of interest, usually for the extraction of epochs.

signal averaging The combination of data from multiple instances of the same manipulation in order to improve functional SNR.

event-related potentials (ERPs) Small electrical changes in the brain that are associated with sensory or cognitive events.

activity of interest will occur for short and discrete intervals, as when a brief flash of light evokes transient activity in the visual cortex. Stimuli that generate such short bursts of neural activity are known as events; in many experiments, a single trial may comprise more than one event. For example, an experiment looking at selective attention may use trials with two events: an initial attention-directing cue followed by a target. In most event-related designs, different conditions of the IV are associated with different events, as in the situation shown in Figure 9.11. Each event is separated in time from the previous event, with an interstimulus interval, or ISI, that can range from about 2 s to 20 s depending on the goals of the experiment. This differs from typical blocked designs, which may present many stimuli consecutively within a task block. Also unlike blocked designs, the different conditions are usually presented in a random order rather than an alternating pattern. Event-related designs have sometimes been called single-trial designs, to emphasize that stimuli are presented one at a time rather than within a block of trials. However, as data collection and analysis procedures have improved, analyses of changes in BOLD signal following only one stimulus presentation have become possible, and thus the label single-trial should be reserved for such experiments.

Event-related designs were rarely used in the early years of fMRI. Most research involved long-interval blocked designs, with the notable exception of the 1992 study by Blamire and colleagues discussed in Chapter 7. Furthermore, even the few studies that involved measurement of the BOLD response to short-duration stimuli did not include the additional analyses (i.e., trial averaging, latency measurements) that event-related designs afford. Within a few years, however, fMRI researchers began to use design ideas from electrophysiology in addition to the concepts drawn from PET. Since the first recordings of electrical activity in the human brain by Hans Berger in the 1920s, researchers had known of tonic changes in the electroencephalogram (EEG) associated with different states of arousal or alertness. These changes were identified by comparing the EEG pattern during one state (e.g., deep sleep) with the pattern during another state (e.g., waking). By the late 1950s and early 1960s, researchers began investigating whether signals associated with specific sensory or cognitive events could be identified within the continuous EEG. By synchronizing or time-locking the EEG signal to the onset of a stimulus and signal averaging across many trials, they could extract small electrical changes known as event-related potentials, or ERPs, from the continuous EEG. Some ERP's, particularly those with short latencies (e.g., <100 ms), were associated with sensory pro-
cessing. Others were associated with cognitive events. For example, when subjects must respond to the appearance of an unexpected stimulus, there arises a systematic positive ERP deflection (now known as the P300) about 300 ms following the stimulus onset. These key concepts, namely time-locking and signal averaging, were critical aspects of early event-related studies.

One of the first event-related studies, conducted by Buckner and colleagues, compared blocked and event-related versions of a word-generation task. Their results were encouraging; with both approaches they identified similar sets of activated regions in the visual, motor, and left prefrontal cortices (Figure 9.12A), although the amplitude of the activation was generally smaller in response to the events (<1%) than to the blocks (2–3%). Moreover, there was a suggestion that the relative timing of the hemodynamic response could be used to identify timing differences between regions (Figure 9.12B and C). In another very early event-related experiment, McCarthy and colleagues examined brain regions associated with the detection of an unexpected and infrequent stimulus. By its nature, this could not be studied within a blocked design, because blocking the stimuli would render them neither unexpected nor infrequent. The authors found that detection of an infrequent target evoked BOLD activation in the dorsolateral prefrontal cortex, chiefly in the right hemisphere, and in the lateral parietal cortex bilaterally (Figure 9.13A–D). Furthermore, they extracted segments from the fMRI time series that were time-locked to the presentation of the events of interest; these are known as epochs, and the result of averaging all the epochs from one condition is an averaged epoch.

**Figure 9.12** Results from one of the first comparisons of blocked and event-related designs. Subjects were shown word stems (e.g., “ora-“) and then generated complete words (e.g., “orange”). The researchers compared blocked and event-related versions of the same design and found generally similar patterns of activation (A), although more activation was observed in the blocked designs. Of additional interest was the suggestion that the event-related design could be used to identify latency differences between brain regions. Parts (B) and (C) each show plots of the responses over time in two regions of the brain, the extrastriate cortex (blue line) and the left prefrontal cortex (red line), from two of their subjects. In Subject 5 (B) the response in the prefrontal cortex was slightly delayed compared with the response in the extrastriate region, although this result was not as clear in Subject 6 (C). (A from Buckner et al., 1996; B, C after Buckner et al., 1996.)

**epoch** A time segment extracted from a larger series of images, usually corresponding to the period in time surrounding an event of interest.

**averaged epoch** The result of averaging a large number of epochs that are time-locked to similar events.
Figure 9.13 An early use of event-related fMRI to study detection of rare stimuli. Subjects watched a series of letters, most of which were “O”s, and pressed a button whenever a rare “X” target was presented. Data were collected from four coronal slices, two within the frontal lobe and two within the parietal lobe. The target stimuli evoked transient activation in the dorsolateral prefrontal cortex (A and B) and in the lateral parietal cortex (C and D), as indicated by the red color overlaid upon activated voxels. Shown are the event-related responses evoked by these target stimuli in the prefrontal (E) and parietal (F) cortices. (A–D from McCarthy et al., 1997; E, F after McCarthy et al., 1997.)

Plotted in Figure 9.13 E and F are averaged epochs containing the target-evoked hemodynamic responses, which exhibit a shape similar to that introduced in Chapter 7.

**Principles of event-related fMRI**

The first event-related experiments provided a new way to think about fMRI data. When stimuli were presented in a blocked design, the resulting BOLD signal was considered to reflect steady-state brain activation at any moment in time. Event-related studies, in contrast, measured transient changes in brain activation associated with discrete stimuli. The pattern of changes over time became critical for experimental analyses. For this reason, high temporal resolution is more important for event-related studies than for blocked studies.
Often in event-related studies, successive images are acquired with TRs of 1 s to 2 s, in order to sample the hemodynamic response at a sufficiently fast rate. Estimation of the time course of the hemodynamic response is often very good, especially when events are presented in relative isolation or when sophisticated strategies are applied to separate the responses to closely spaced events.

From the linear systems framework that was introduced in Chapter 7, one can consider stimulus events as impulses, each of which evokes a hemodynamic response. As first demonstrated by Boynton and colleagues in 1996, the amplitude and timing of a hemodynamic response depend on both the intensity and the duration of the evoking stimulus. To the extent that the linear assumption holds (i.e., that the BOLD responses to successive stimuli do not interact), it is possible in principle to present events very rapidly and extract the hemodynamic response associated with individual events. Dale and Buckner investigated this issue in a 1997 study. They hypothesized that signal averaging could be applied to event-related designs even if the events were presented only a few seconds apart, despite the much slower (10–15 s) rise and fall of the hemodynamic response. The stimuli were flashing checkerboards of one-second duration presented to either the left visual field or the right visual field (remember that a visual stimulus presented to the left evokes activation in the right visual cortex, and vice versa). Importantly, the order of presentation of the stimuli was randomized, not alternated, because alternation at such short interstimulus intervals would have removed any effects (see Figure 9.9, for example).

**Thought Question**

How would the assumptions of event-related fMRI be violated if the neuronal activity in response to an event was not an impulse but instead had a long duration (e.g., 5 s)? How would the hemodynamic response change?

Figure 9.14 shows data from one of their experiments. At interstimulus intervals of only 2 s, robust activation for each trial type could be detected in the contralateral primary visual cortex. In fact, the observed activation was more easily detected at shorter intervals than at longer intervals. This seems counterintuitive, but remember that experimental power depends on the amount of data that is averaged. In the 2-s condition, there were many more trials per experimental run than in the longer-interval conditions. When more trials are presented in rapid succession, the total variance in the BOLD signal increases, resulting in more experimental power. Dale and Buckner demonstrated that, provided the events of interest are presented in a random order, areas of BOLD activation can be detected even using very short interstimulus intervals.

Now, does this study imply that short intervals are best for event-related studies? Not necessarily, because these results focused on the detection of areas of activation, not estimation of the time course of the hemodynamic response. For accurate time-course estimation, many researchers have used periodic event-related designs that presented the events of interest at regular intervals. Slow (more than 1 s ISI) periodic designs are conceptually very simple, and their analysis is straightforward. Each event evokes a complete hemodynamic response, and events can be combined through selective averaging. Slow periodic designs are inefficient, however, due to their low frequency of events over time. Fast periodic designs would seemingly be more efficient, following the
Figure 9.14 Rapid event-related fMRI with randomized stimulus presentation. Functional MRI activation associated with different stimulus types can be extracted from very rapid event-related designs, provided that the order of the stimuli is randomized. In this experiment, flickering visual checkerboard stimuli were randomly presented to either the left or right visual field with 2-, 5-, or 10-second inter-trial intervals. Increased activation in the correct hemisphere (i.e., left visual cortex for stimuli presented to the right hemifield) was observed at all intervals, as shown here for a single subject. The bar scale at right indicates the significance values associated with each color on the images. Interestingly, the lowest significance values and smallest spatial extent of activation were found at the longest inter-trial interval, due to the much reduced number of experimental trials that were presented there compared to the shorter-interval conditions. (From Dale and Buckner, 1997.)

logic of the previous paragraph, but in fact may be even less practical. In 2000, Bandettini and Cox attempted to determine the best interstimulus interval for periodic presentation of a short-duration (2-s) stimulus. They found that at intermediate intervals of 10–12 s, substantial stimulus-related variability was present in the data, as shown in Figure 9.15A and B. However, as they shortened the interstimulus interval, the effects of the individual trials became less and less apparent. At the shortest interval tested, 2 s, there was an increase in BOLD activation at the beginning of the run but a plateau thereafter, precluding analysis of any individual trials.

Thus, a short-interval periodic design will cause the BOLD signal to saturate to some maximum value, so that the identification of stimulus-related effects is not possible, making the experiment essentially worthless. Many researchers now use experimental designs that involve short, but variable, intervals between successive stimuli, either by jittering the time interval between successive stimuli or by randomizing the order of different events. Based on data from a number of published studies, some using simulations and others empirical trials, there is now good evidence that the best experimental designs use variable interstimulus intervals in which successive events of the same type occur, on average, every 4 s to 6 s. (Note that this mean inter-
Figure 9.15 Effects of interstimulus interval on event-related fMRI activation. As the interval between successive events decreases, the overlap between consecutive hemodynamic responses reduces the variability in the BOLD signal. Subjects performed a finger tapping task while watching a flashing visual stimulus. Activations within regions of interest in the visual cortex (A) and motor cortex (B) were measured under a number of different experimental conditions. When there was a long interstimulus interval (ISI) of 20 s and a long stimulus duration (SD) of 20 s, mimicking a blocked design, there was clear alternating activation in both regions. However, for short-duration events of 2 s, periodic activation was present at long ISIs of 10 s to 12 s but not at short ISIs. (From Bandettini and Cox, 2000.)

Val is sufficiently long to allow near-full recovery from the refractory effects described in Chapter 7.) Box 9.2 describes the key concepts of optimal experimental design in much greater detail.

We emphasize that the timing constraints described here apply to the interval between successive events and not necessarily to the interval between successive stimuli. Imagine an experiment that presented two types of stimuli, words and nonword letter strings. If you knew that a given brain region responded to the words but not the non-words, you could embed the words in a rapidly presented series of nonwords and the words would still evoke a large hemodynamic response. For that brain region, only the words would serve as events. However, for another brain region that responds equally to both words and non-words, both types of stimuli would serve as events and BOLD activation would be at a steady state throughout the experiment. This logic is identical to that used in the study by McCarthy and colleagues described in the last section, and bears similarities to the fMRI-adaptation paradigms described in Chapter 8.

Advantages of event-related designs

Event-related designs have become increasingly common, now representing the lion’s share of all fMRI research. This popularity has resulted from several advantages, as outlined here, combined with the increasing flexibility of fMRI analysis programs for handling the statistical models resulting from very complex designs.
We have emphasized throughout this chapter that different types of designs have distinct advantages and disadvantages. Given the wide variety of questions that can be asked using fMRI, there is no way to create the perfect experimental design. However, there are many, many ways to create an imperfect, inefficient, or just plain bad experimental design.

What makes an experimental design efficient or inefficient? Recall that the overall goal of experimental design is to allow the researcher to test some research hypothesis. Most such hypotheses, if well-formed, involve differentiating between two or more experimental conditions. The efficiency of a design depends on its ability to distinguish between those conditions in terms of their levels of activation; efficient designs allow conditions to be differentiated using a relatively small fMRI dataset, whereas an inefficient design would require much more data to reach the same conclusion. One common problem of inefficient experimental designs is a high degree of correlation between the processes of interest. Correlation could arise because processes frequently co-occur (e.g., whenever Event A happens, Event B happens as well) or because distinct processes follow one another too closely in time (e.g., a cue to direct attention always occurs 100 ms before the target stimulus). Designs will also be inefficient if they are poor at evoking the processes of interest (and thus do not generate maximal BOLD signal changes). This can be particularly problematic when complex processes, or subjective states like aggression, regret, or decision making, are being studied. If subjects fail to take the task seriously, or engage in processing not anticipated by the experimenter, then an otherwise good design may become worthless. Finally, a design could be inefficient simply because it is poorly matched to the research hypothesis; a canonical example would be using a pure blocked design, but attempting to estimate the shape and timing of the hemodynamic response evoked by each stimulus within the blocks. Note that the same blocked design might be very efficient for answering a different sort of research question.

Two critical concepts are worth re-emphasizing: efficiency depends on the specific experimental hypothesis, and it is measured by the amount of data that would be required to test that hypothesis. These concepts guide researchers who study how to create more efficient experimental designs. Broadly considered, these researchers ask the question, “Given a particular sort of hypothesis to be tested, how should I present experimental stimuli to maximize the size of my effects?” Key factors that have been studied include the ideal interval between successive experimental stimuli (whether events or blocks), whether that interval should be fixed or variable, and whether and how to randomize the order of stimuli. We will consider each of these factors in turn. For clarity, our examples will describe designs involving a single stimulus type (e.g., “what activation is evoked by the presentation of a face?”) rather than two or more stimulus types (e.g., “what differential activation is evoked by faces vs. houses?”). Note, however, that the logic of the following argument holds for all types of designs.

The fMRI activation evoked by a single stimulus can persist for 15 s or more, considering the rise, fall, and undershoot of the BOLD hemodynamic response. Blocked designs take advantage of this sluggishness by effectively integrating activation across a number of successive stimuli. If the duration of the blocks is too short—less than about the duration of the hemodynamic response—then the differences between conditions will not be maximized. Remember that too-long blocks also have a disadvantage: if the blocks alternate at a relatively low frequency, then there will be greater noise at the frequency of alternation. For this reason, the optimal block length in an fMRI experiment is often around 20 s. A natural inference is that similar constraints apply to event-related designs. Many early researchers spaced successive events widely, with intervals of about 20 s, so that the response evoked by one stimulus does not overlap with the response evoked by another stimulus. For regularly-spaced events (i.e., periodic designs), intervals longer than about 15 s are indeed necessary for estimating the full shape of the hemodynamic response (see Figures 9.9 and 9.15 for examples). Yet, as shown by Dale in a 1999 article, this advantage completely disappears if the interval between successive stimuli (of the same type) is variable. For example, a periodic design with a fixed interstimulus interval of 4 s is extremely inefficient, because the BOLD signal quickly rises to its maximum value and remains high for the duration of the experiment. However, if the same stimuli are presented with interstimulus intervals that vary around a mean of 4 s, the ability to detect a significant BOLD effect (and even to estimate the shape of the hemodynamic response) increases remarkably! Subsequent work has revealed that the most efficient experimental designs are often event-related with very short, but highly variable, intervals between successive stimuli (see Figure 1).
BOX 9.2 (continued)

Figure 1 Efficiency of variable-interval and fixed-interval event-related designs. At relatively long intervals between stimuli (i.e., those greater than the typical width of the hemodynamic response), whether that interval is variable or fixed makes relatively little difference—both are relatively inefficient. However, at shorter intervals there is a marked increase in the efficiency of designs that incorporate variable intervals between successive stimuli. In contrast, if intervals between successive stimuli are short, the efficiency of a fixed-interval design (i.e., a periodic design) declines to zero. Designs that use fixed, but short, intervals have essentially no power, as is also shown in Figure 9.15. (After Dale, 1999.)

Why does using a variable, short interstimulus interval so dramatically increase the efficiency of fMRI experimental designs? The answer can be found by considering the properties of the time course of activation evoked by these designs. (We describe these properties conceptually in this section and refer those interested in the underlying mathematics to the papers by Dale, Liu, Wager, and their colleagues listed in the Chapter References.) When a few events occur in rapid succession, the combined hemodynamic response will increase to a near-maximal value. But, when events fail to occur for some time, the hemodynamic response will return to near-baseline levels. Thus, activated voxels will undergo large changes in amplitude over time, and the design has relatively good efficiency for research hypotheses involving the detection of those voxels. Moreover, variable intervals can have a second useful property: successive time points are relatively uncorrelated. Sometimes consecutive events will immediately follow each other, sometimes there will be a short gap between events, and sometimes there will be a long gap. To the extent that the presentation of one stimulus does not predict the occurrence of the next stimulus, this design has a maximal ability to estimate the shape of the hemodynamic response evoked by those events. Thus, introducing variable intervals between successive events can maximize estimation efficiency while maintaining very good detection efficiency. Another advantage of variable intervals is that they minimize the subjects' ability to predict when (and what) events will occur. This can reduce psychological processes associated with anticipation and expectation, while also preventing subjects from using simple strategies to perform the experimental tasks.

So far, we have emphasized the value of an efficient experimental design, while glossing over exactly how to construct one! Several approaches have been used in published research. Most common is simple randomization. If there are two event types that could occur with equal probability in a random sequence, the same event could occur several times in a row, or one event could not occur at all in a sequence of many stimuli. This results in a distribution of intervals that approximates that of an exponential decrease. One can also jitter the relative time between successive stimuli, as is necessary for short-interval designs with only one stimulus type, by determining the interstimulus interval using random draws from some distribution of possible intervals (e.g., a uniform distribution comprising integral values between 1 s and 7 s). Another approach is to have the statistical properties of the experimental design actually change over time. In a semirandom design (Figure 2), stimuli are sometimes likely to occur and sometimes unlikely to occur. For example, imagine that your event is a flashing checkerboard presented for 500 ms, and you are sampling the brain at a semirandom design. A type of event-related design in which the probability that an event will occur within a given time interval changes systematically over the course of the experiment.

(continued on next page)
BOX 9.2 (continued)

Figure 2. Semirandom designs combine features of blocked and event-related designs. The vertical lines below each graph represent the timing of each single stimulus, such as a flash of a visual checkerboard, and the graphed curves reflect the expected hemodynamic response. The semirandom design contains large-scale structure, in that some time intervals have high event-probability and some have low event-probability, with small-scale randomness. (A) and (B) combine a blocked epoch with random and semirandom periods, respectively, while (C) is a completely semirandom design. Note that these three designs have equal estimation efficiencies and detection powers. (After Liu et al., 2001.)

TR of 1000 ms. You can set up your experiment so that in the first 30 s block, each TR has a 25% chance of containing an event; in the next 60 s block, each TR has a 75% chance; and in the final 30 s, each TR has a 25% chance. The resulting design would look similar to Figure 2C. Although composed of individual events, semirandom designs are similar to blocked designs in that some time periods contain many events while others contain very few; this clustering increases the total variability in the BOLD signal. Work by Liu and colleagues has shown that, consistent with the conceptual predictions in the previous paragraph, semirandom designs can have detection power that are almost as good as those of blocked designs while also having optimal estimation powers.

By their nature, the properties of randomized experimental designs are... random. A random process might generate a sequence that has the virtues described in this box (i.e., event clustering, lack of temporal correlations), or it might not. As a simplified example, suppose that your experiment involved two stimulus types, here labeled “0” and “1,” that were equally likely to occur over ten equally-spaced trials. A random process like a series of coin flips could generate any of $2^{10}$ possible sequences (e.g., 1010011101, 1111111100, or 0101010101) that vary considerably in their efficiency. Typical fMRI experiments contain tens if not hundreds of events, and the intervals between those events are rarely completely constrained by the cognitive demands of the task. Given such a multitude of possible stimulus orders and timings, the odds of simple randomization providing an experimental design that is near-optimal are vanishingly small. Efficiency can be improved significantly by generating many different random designs and then selecting the best alternative (i.e., performing a random search), but even this is unlikely to provide a near-optimal design, especially given the complexity of most fMRI experiments.

Researchers have therefore focused on developing algorithms that can quickly evaluate parameters of possible designs to identify the most efficient. Some researchers have used the tools of mathematics to generate rules for counterbalancing the order of events, leading to designs that are very good at estimating the hemodynamic response evoked by each stimulus class (e.g., the work of
Buracas and Boynton and of Aguirre cited in the Chapter References). One particularly exciting approach, developed by Wager and Nichols, is the use of genetic algorithms. Like their biological namesakes, genetic algorithms evaluate the relative fitness (i.e., design efficiency) of each design, which consists of many smaller units (i.e., the timing of events). Then, some aspects of the relatively good designs are combined (i.e., some events are swapped), while some elements of the designs undergo mutation (i.e., they change randomly), and then the new designs are tested for their levels of efficiency (Figure 3). Note that the least efficient designs are discarded, and thus options that are unlikely to lead to optimal designs are not explored. This process iterates until it reaches some criterion, such as obtaining a sufficiently good design, or reaching some time limit for searching. Simulations by these authors, using genetic algorithms, resulted in the rapid identification of designs that were near-optimal for specific criteria (e.g., that maximized estimation efficiency while retaining good detection power). Moreover, these designs were demonstrably better than those obtained by randomization or even random searches.

It is important to emphasize that, just as no single experimental design is appropriate for all research hypotheses, no approach for optimizing experimental designs is without specific limitations. Consider two tasks, one involving the passive viewing of discrete visual stimuli (e.g., 500 ms checkerboards), and the other involving making decisions about complex moral dilemmas. Algorithmic criteria that lead to an optimal design in the former case may select a very poor design for the latter case, simply because of the influence of time pressure on the way people evaluate complex decision scenarios. Returning to our cardinal rule for fMRI design, the most important thing to consider is whether the experimental design evokes the mental processes—and thus the neuronal activity and, in turn, the BOLD signal—that you intended.
trial sorting The post hoc assignment of events to conditions, often based on behavioral data.

subsequent memory An approach to fMRI analyses that sorts experimental stimuli based on whether they were remembered or forgotten in a later testing session; this allows identification of brain regions whose activation predicts successful encoding of stimulus properties into memory.

Estimations of the shape and timing of the hemodynamic response are generally much better for event-related designs than for blocked designs. Estimation is important for many types of research questions. By characterizing the precise timing and waveform of the hemodynamic response, researchers can make inferences about the relative timing of neuronal activity, about distinct processes within separate parts of a trial, about functional connectivity between regions, and about sustained activation within a region. Conversely, blocked designs generally have superior detection power because events are concentrated within the task blocks, whereas simple event-related designs evoke smaller overall changes in the BOLD signal. More complex event-related designs, however, can be optimized so that they can have very good detection power without sacrificing estimation efficiency (see Box 9.2), ameliorating the major disadvantage of event-related designs.

Even so, decisions about experimental design cannot be made solely on the basis of the relative importance of detection versus estimation. The major goal of fMRI research is understanding brain function, and so experiments should be designed so that specific functions can be readily individuated (e.g., separated in time). In this regard, event-related designs provide a degree of flexibility not present in most blocked designs, as the same events can be analyzed in different ways depending on the goals of the experiment. A researcher interested in how the presentation of an image affects its perception might present two types of images (e.g., faces and objects) in two orientations (e.g., right side up and upside down), with the images presented one at a time and in random order. These stimuli could be considered as one type of event, visual stimuli; two types of events, faces or objects; or even four distinct types.

Note that this flexibility means that researchers can even choose their events based on experimental data, often known as trial sorting. Like many of the characteristics of event-related fMRI analyses, the basic concepts of trial sorting are derived from earlier electrophysiological studies. Trials are often sorted based on subject responses, like accuracy or response time. For example, some brain structures, such as the anterior cingulate gyrus, are more active during trials in which the subject makes an error than during trials when the subject does not make errors. These differences may reflect cognitive processes like recognizing the mistake, adjusting response plans to prevent future errors, or reflecting on the cause of an error. Trials can even be sorted by their long-term consequences. In experiments using the subsequent memory paradigm, subjects are presented with a large series of items (e.g., pictures of unfamiliar scenes) while within the fMRI scanner. Usually, but not always, they perform some task unrelated to memory (e.g., judging whether the scene is indoor or outdoor). At some later time, subjects complete a memory test that includes these items and others that were not actually presented. Then, the researchers sort the original stimuli into categories of “remembered” and “forgotten,” to identify brain regions associated with successful encoding of information into memory. The first such studies were reported in 1998 by Brewer and colleagues and by Wagner and colleagues, and since then, researchers have used this approach to map out the contributions of many regions to memory. For example, in a 2006 study by Adcock and colleagues, subjects were presented with a reward cue (e.g., $5 or 10¢) before each item to be remembered, so that the subjects knew how much money they could win if they successfully remembered that item. They found that for those items that were later remembered, there was increased activation in brain regions associated with evaluation of rewards (e.g., the striatum and the ventral tegmental area within the midbrain), along with increased correlations between those regions and the hippocampus. Conversely, memory impairments resulting from prior sleep deprivation,
as reported by Yoo and colleagues in 2007, are associated with reduced subsequent memory effects in the hippocampus, along with changes in the functional connectivity of that region.

Because of these advantages in flexibility, a large and increasing proportion of fMRI experiments have adopted event-related designs. But as we indicated earlier, event-related designs generally have lower detection power than similar blocked designs, resulting in part from their sensitivity to the shape of the hemodynamic response. If the wrong model for the fMRI hemodynamic response is used, then significant activations may be missed. Nevertheless, for a wide range of experimental tasks, event-related designs provide the best combination of flexibility and experimental power.

Mixed Designs

Some fMRI experiments combine the basic elements of blocked and event-related approaches. Within a mixed design, stimuli are presented in discrete and regular blocks, but within each block are multiple types of events. This is illustrated in Figure 9.16. An important difference between mixed designs and the other types of designs is that mixed designs allow analysis of IVs that change on different time scales. The task blocks, which may last for 20–30 s or more, are associated with sustained changes in task strategy, attention, or other cognitive processes. A subject may be attending to the left side of the visual display during one block and then attending to the right side for the next block. The different blocks induce different cognitive states in the subject, and thus blocked analyses can measure state-related processes. This differs conceptually from the semirandom designs described in Box 9.2, in that mixed designs assume that the grouping of events into a task block will cause the subject to adopt and maintain a particular cognitive state, whereas the individual stimuli in a semirandom design are assumed to evoke a particular cognitive process repeatedly (i.e., no underlying cognitive state emerges). Thus, mixed designs are appropriate when one wants to examine sustained brain activation, while optimized event-related designs are preferred for detection of transient brain activation.

As in an event-related design, the stimuli presented rapidly within the block will evoke separable short-term changes in the brain. For example, a subject may have been directed to watch a visual display on which stimuli will appear repeatedly on both the left and right sides, but to attend specifically to the left side on some blocks and to the right side on other blocks. Different events will evoke different processes; for example, an object presented on the attended left

Figure 9.16 An example of a mixed fMRI design. In a mixed design, events of interest are clustered within extended blocks. Here, the task blocks involved an “oddball” paradigm in which subjects pressed buttons to the occurrence of infrequent target circles and ignored the non-target squares. Each circle was thus an event that evoked item-related cognitive processes. The task blocks alternated with non-task blocks (red triangles) that required no responses. Thus, the difference between the task and non-task blocks reflected state-related processes. (After Huettel et al., 2004.)
item-related processes. Changes in the brain that are assumed to be caused by the properties of individual stimuli, or items. Item-related processes are more easily measured with event-related designs.

will lead to different changes in brain activation than an object presented on the unattended right. The individual events within a mixed design’s blocks reflect item-related processes. Note that state- and item-related processes are not necessarily related within a task; for example, the brain structures responsible for attending to the left or right are not the same as those responsible for pressing buttons in response to targets. Mixed designs are thus extremely powerful for research questions that involve both a long-term cognitive process and short-term implementations of cognitive processes during individual trials.

Thought Question

Mixed designs can also be used to study transient processes that occur at the onset or end of a block. What sorts of cognitive processes are likely to occur at the beginning and end of a task block?

One of the first studies to demonstrate the power of mixed designs was reported in 2001 by Donaldson and colleagues, who wanted to differentiate between brain regions associated with attempted memory retrieval and regions associated with retrieval success. Their experiment consisted of long (105 s) task blocks that each contained 42 events of three types, presented in random order: words that had been studied in a prior practice session, words that had not been studied previously, and fixation events in which no word was presented. Sustained increases in activation during the task blocks were observed in the prefrontal and insular cortices, whereas event-related activation was found in a more diverse set of regions that included the visual cortex, the thalamus, and other parts of the prefrontal cortex. Only by using this sort of mixed design could the authors separate brain systems that underlie state- and item-related aspects of their task. As another example, a 2006 study by Dosenbach and colleagues used mixed designs in experiments involving ten different fMRI tasks and 183 subjects. Their interest was in characterizing brain regions that exhibited both increased activation at the onset of a task block and sustained activation throughout the block, regardless of the particular task that was being performed. Two regions, the anterior insula and anterior cingulate cortex, met these conditions, while also responding specifically to error trials. The authors’ conclusion, that these regions constitute a system for establishing and maintaining task sets, could not have been reached without using a mixed design.

It is important to recognize, as noted by Donaldson and colleagues in their article, that state-related activation does not necessarily result in increased activation during one block type compared with another block type. Changes in a cognitive state could instead have modulatory effects on event-related activation, so that the hemodynamic responses to individual events within the block increase in amplitude. Consider a design in which task blocks involve attention to stimuli while nontask blocks do not. Within a region associated with the control of attention, there might be increased, steady-state activation throughout the task block. But in a region whose activation is influenced by attention, there might be no state-related effects, but instead larger hemodynamic responses to events within attended blocks than within unattended blocks. Mixed designs thus have great utility for many types of cognitive questions, but require analysis strategies targeted to the expected activation of interest.
Summary

The important issues in fMRI experimental design are the creation of a research hypothesis, the choice of experimental conditions to test that hypothesis, and the presentation of stimuli that manipulate the experimental conditions over time. When selecting conditions for an experiment, it is important to avoid confounding factors, or variables that unintentionally covary with the independent variable of interest. There are three main types of fMRI experimental designs: blocked, event-related, and mixed designs. In a blocked design, each condition is presented continuously for an extended period (one block) and the different blocks of conditions are usually alternated. Some blocked-design studies incorporate a baseline period without an experimental task, to account for specific brain regions that are often more active in baseline conditions. Event-related designs present stimuli one at a time rather than together in a block. Long-interval periodic event-related designs are useful when a prestimulus baseline is necessary but otherwise have poor experimental power. More desirable are event-related designs that use variable but short intervals between successive events of interest to optimize experimental power. Mixed designs combine blocked and event-related analyses and are used to distinguish between long-term sustained activation and short-term transient activation.

There is no perfect experimental design for all fMRI studies. The fundamental rule when designing your study is that you should choose the design that best suits your experimental question. Even a seemingly optimal event-related design could be useless if it does not evoke the cognitive processes of interest, just as a blocked design cannot be used if the conditions cannot be separated into different blocks. When designing a research study, the most important factor to consider is the simplest: which design will allow my experimental manipulation to evoke measurable differences in BOLD activation?

Suggested Readings


This article combines simulation and experimental work to demonstrate the effectiveness of genetic algorithms for identifying optimal parameters and stimulus presentations for fast event-related fMRI designs.
*Indicates a reference that is a suggested reading in the field and is also cited in this chapter.

**Chapter References**


