2

# Research Design and the Behavioral Analysis of Drug Effects

## RESEARCH DESIGN

All scientific experimentation can be thought of as a search for a relationship between events. In behavioral pharmacology, the researcher is usually trying to find the relationship between the presence of a drug in an organism and changes that occur in the behavior of that organism. In most true experiments, one of these events is created or manipulated by the experimenter, and the other event is measured. The manipulated event is called the *independent variable*, and the observed event is called the *dependent variable*. The independent variable in behavioral pharmacology is usually the amount of drug put into the organism; that is what the researcher manipulates. The dependent variable is usually some change in the behavior of that organism, and this is what the researcher measures. Later in this chapter we will discuss some of the more commonly used measures of behavior or dependent variables.

### Experimental Research Design

***Experimental Control.*** It is not enough to give a drug and observe its effect. For an experiment to be meaningful, the experimenter must be able to compare what happened when the drug was given with what would have happened if the drug had not been given.

A controlled experiment is one in which it is possible to say with some degree of certainty what would have happened if the drug had not been given. This permits comparisons between drug and nondrug states. For example, a researcher could give each person in a group of subjects a pill containing THC, the active ingredient in marijuana, and observe that everyone tended to laugh a great deal afterward. These observations would not be worth much unless the research could demonstrate that the increased laughter was a result of the drug and not a result of the subjects' expectations, their nervousness about being observed, or some factor other than the presence of the drug in their bodies. With most behavioral experiments, many factors could influence the results, so it is essential to be sure that the drug and not something else in the procedure caused the laughter.

The only truly reliable way to do this experiment and eliminate all other possible causes of the laughter would be to have a time machine and, after the experiment, go back and, under exactly

the same circumstances, give the same subjects pills identical in appearance but not containing any drug. Comparisons could then be made between the amount of laughter with and without the drug because all other factors (the subjects, the situation, the time of day, and so on) would be the same. Only then could we be sure the laughter was caused by the drug and nothing else.

There is no such thing as a time machine, so the behavioral pharmacologist must compare the behavior of a drugged subject with either (a) the drug-free behavior of that subject under similar conditions (within-subject design) or (b) the behavior of other drug-free subjects under similar conditions (between-subject design).

***Within-Subject Design.*** In this strategy, careful observations are made of a subject's behavior under specific conditions. When the behavior appears to be stable and predictable, it is then possible to give the drug and make comparisons between drugged and nondrugged behavior. In other words, subjects serve as their own controls.

Let us say, for example, that we are interested in the effect of amphetamine on the feeding behavior of rats. In a within-subject design, the researcher carefully measures the daily food consumed by several rats until the measures are constant for each animal. The researcher then injects a dose of amphetamine into each rat and measures food consumption for that day. A meaningful comparison can then be made between food consumption with the drug and without the drug because the researcher has a pretty good idea of how much each animal would have eaten if it had not been given the drug.

***Between-Subject Design.*** The same experiment could be done using a between-subject design. In this strategy, a number of rats are randomly assigned to two groups. One group, the experimental group, would get the amphetamine. The other group, the control group, would not get any drug. The food consumption of both groups could then be compared.

***Comparisons of Between- and Within-Subject Designs.*** The type of design used by a behavioral pharmacologist is usually determined by the type of dependent variable being measured in the experiment. If the measured activity (eating, in this case) is stable from day to day, within-subject designs can be used, but if the dependent variable is subject to systematic change, the researcher is forced to use a between-subject design. Exploratory behavior is a good example of such a measure; on the first exposure to a new cage, a rat will usually spend considerable time moving around and exploring, but on the second day, it may be habituated to the surroundings and may just sit and lick its whiskers. A within-subject design could not be used to study exploratory behavior because the behavior changes from day to day. We would not know whether the change was due to the drug or to habituation. A between-subject design would be appropriate because both the experimental and control groups could be compared on the first exposure to the new cage, and habituation would not be a factor.

The difficulty with the between-subject design is that responses of individuals may vary a great deal. By chance, there may be very curious rats in one group and very lazy rats in another. The differences in the groups would then be due to differences in the rats and not in the effects of the drug. We could get around this difficulty by having larger groups to decrease the likelihood that all the curious rats (or all the lazy ones) would end up in one group. This procedure would require much more work and would have the additional disadvantage that the final results would be in terms of group averages, which sometimes hide important information that is more apparent when working with individual subjects.

The advantage of the within-subject design is that more perfect control conditions can be achieved because each subject is its own control. The disadvantage is that it can be used only with behavioral measures that are not likely to change when repeated. Within-subject experiments usually take more time, but they do not require as many subjects.

### Statistical Testing

In some within-subject and most between-subject designs, some statistical tests are needed to determine the probability of differences observed between drug and nondrug measures. Such tests are necessary because differences could be due to chance variations from day to day and from subject to subject. When a researcher finds a difference in the means of the groups, a statistical test can tell how often such a difference would be likely to occur by chance if there were no drug effect. Box 2-1 gives a numerical example of how statistical tests are used.

### Placebo Controls

To be completely useful, a control condition must be as similar as possible to the experimental condition, except for one variable: the presence or absence of the drug. In our example in which the effect of amphetamine on rats' eating was determined, the control procedure could have been improved. As you recall, we had two groups: One was injected with amphetamine, and the other was not injected at all. It is quite possible that the anxiety of being stuck with a needle suppressed eating and that the amphetamine had nothing to do with the results. For this reason, behavioral pharmacologists always use a control condition that involves the injection of the vehicle alone (see Chapter 1), that is, a *placebo control*. On control days in within-subject designs and for control subjects in between-subject designs, an injection of normal saline would be given. Subjects in the experimental group or those receiving amphetamine on drug days would be treated identically, except that they would have the drug dissolved in the saline.

### Placebo Effects

Such careful controls are especially important with human subjects because of the *placebo effect* phenomenon. A placebo is a totally inert substance that causes no physiological change but is administered as though it were a drug. If people believe they are getting a drug that will have a specific effect, they will frequently show that effect even though the drug does not cause it.

In an interesting experiment by Fillmore and Vogel-Sprott (1992), three groups of subjects were given a cup of coffee before being tested on a psychomotor performance task. One group was told that caffeine would speed their performance, one group was told that caffeine would slow their responding, and the third group was not told anything. The coffee all were given was a placebo: decaffeinated coffee. The groups' performances matched what the subjects had expected; those who were told to expect improvement did better than those who were not told anything, and those told to expect an impairment did worse than those who were not told anything.

This placebo effect makes careful control an absolute necessity when evaluating the clinical effectiveness of newly developed medicines because patients will frequently show an effect they expect the drug to produce. For example, let us suppose that we are testing a new pain reliever. We go to a hospital and give the drug to a group of patients who are in postoperative pain and tell them that this new drug should relieve their distress. The next day, we find that 68 percent of the patients report that their pain was relieved. By itself, this is not a useful experiment because we do not know how many patients would have reported the same thing without the drug. To do this experiment the proper way, it would be necessary to have two groups of patients. Both groups would be told they were getting a pain reliever, but only one group would get the new drug; the other would be given an identical pill containing only an inactive filler. The next day, pain ratings would be taken from all the patients, and comparisons could be made.

The *balanced placebo design* was developed in the mid-1970s by George Marlatt and his colleagues at Washington State University in Seattle (Marlatt & Rohsenow, 1980). It remains the gold standard for research with humans in which the subject's expectations could influence the results in a manner similar to the Fillmore and Vogel-Sprott experiment described previously.

**BOX 2-1** An Example of the Use of a Statistical Test

A mythical experiment was done to determine the effect of amphetamine on food consumption in rats. The experiment was a between-subject design with 10 rats in each group. Both groups were treated identically except that the rats in one group were given amphetamine before eating and the rats in the other group were given a placebo injection of the vehicle, normal saline. The amount eaten by each rat is given in the following table:

| *Control Group* | | *Experimental Group* | |
|---|---|---|---|
| Rat Number | Food Eaten (grams) | Rat Number | Food Eaten (grams) |
| 1 | 18 | 11 | 17 |
| 2 | 22 | 12 | 14 |
| 3 | 23 | 13 | 12 |
| 4 | 17 | 14 | 25 |
| 5 | 28 | 15 | 15 |
| 6 | 20 | 16 | 16 |
| 7 | 16 | 17 | 20 |
| 8 | 22 | 18 | 21 |
| 9 | 21 | 19 | 13 |
| 10 | 18 | 20 | 11 |
| Mean | 20.5 | | 16.4 |

The researcher found that the amphetamine (experimental) group ate a mean of 16.4 grams of food and that the control group ate a mean of 20.5 grams of food. On the basis of this difference, could the experimenter conclude that amphetamine reduced food intake? To answer this question, a statistical test is needed.

The most appropriate statistical test for this type of experiment is what is known as a *t* test. This test takes into account the variability in each group, and, on the basis of certain assumptions, it can tell the experimenter how many times the experiment would have to be repeated to get this big a difference between the means just by chance.

In this case, the *t* test tells the experimenter that if the drug had no effect and the experiment was done 100 times, a difference this big could occur as often as 10 times just by chance. This result is normally not good enough. Most behavioral researchers insist that their results be explained by chance no more than five times in 100, or, as they say, with a probability of less than 5 percent ($p < .05$). If the probability level is greater than .05, the result is generally considered to be negative. If you read original research reports, you will see the results of statistical tests reported something like this: ($t = 2.09$, $df = 18$, $p < .10$). The first two numbers are values associated with the statistical test, and the final number is the probability level (in this case, it is less than 10 percent).

There are many types of statistical tests for many different research designs. However, they all end up telling you the same thing: how often you would expect to get the results by chance if there were no drug effect (Ferguson, 1966).

When using a within-subject strategy, a statistical test is sometimes not necessary. Most research using operant techniques examines the behavior of three or four experimental subjects in great detail and under carefully controlled experimental conditions. Nondrugged performance is very reliable, so when the drug is given, it is readily apparent whether there are any drug-produced changes in behavior, and statistical analyses are not necessary.

In the balanced placebo design, there are four groups: Two are the same, as in a standard placebo design where subjects in both groups are expecting to get a drug and one group gets a drug and the other gets an identical placebo. In the balanced placebo design, there are two other groups. The subjects in neither expect to get the drug, but one gets the drug, and the other group gets a placebo.

This design is powerful at separating the drug effect from the expectancy or placebo effect because there is a group that does not expect the drug and gets it. Any change in this group must be due entirely to the drug. There is also a group that expects the drug and gets a placebo. Any change seen in this group must be due entirely to the expectation effect and not the drug.

In Chapter 3 there is an extended discussion of the nature of the placebo effect.

### Experimenter Bias

Further precautions must be taken in an experiment of this nature. It has been known for some time that an experimenter can influence the outcome of research without knowing it. For example, if the researcher knows which patients have been given a placebo, the researcher might unconsciously change the manner in which the patients are interviewed or even make systematic mistakes in recording data. To eliminate experimenter bias, it is usually necessary to conduct the experiment so that neither the patients nor the researchers giving the drug and interviewing the patients know who has been given the drug and who has been given the placebo. This procedure is called a *double-blind procedure*, and it is essential because it eliminates the possibility of placebo effects and experimenter bias effects.

### Three-Groups Design

When a new drug is being tested for use in the treatment of a disease, the standard design is what is known as a *three-groups design*. One group is given the experimental drug to be tested, a second group is given a placebo, and a third group is given an established drug with known therapeutic effect. By having three groups, the researchers can answer a number of important questions. Comparisons between the experimental drug group and the placebo group show whether the drug caused any improvement, comparisons between the placebo group and the established drug group indicate whether the research measures were sensitive enough to detect an improvement, and comparisons between the experimental and established drug groups tell whether the new drug has any advantage over established treatment (Overall, 1987).

In some cases where denying a patient any treatment is known to cause harm, the placebo group may not be used at all, and new treatments are compared only with old established treatments.

## NONEXPERIMENTAL RESEARCH

A good deal of what we know about drugs is a result of research that does not involve experiments. As explained earlier, experiments attempt to find relationships between two events: a manipulated event and a measured event. *Nonexperimental research* looks for a relationship between two measured events. A good example is the discovery of a relationship between smoking during pregnancy and infant mortality. It was shown some time ago that there was a higher rate of infant death among babies born to women who smoked during pregnancy than among babies born to nonsmoking mothers (see Chapter 9). In this research, nothing was manipulated; there was no independent variable. The two events, smoking and infant mortality, were measured and found to be related.

One major difficulty with this type of finding is that we cannot assume a causal relationship between the two related events. We know that children born to smoking mothers are more likely to die, but we cannot conclude that smoking causes the infants' deaths. The relationship might be due to some third factor that causes both events. For example, it may be that women smoke because

they have a biochemical imbalance that causes their bodies to need the nicotine in cigarettes. This imbalance might also be responsible for the higher infant mortality rates. The only way we could be sure that the smoking caused the infant mortality would be to do a true experiment by finding two groups of nonsmoking pregnant women and forcing one group to smoke. If there were a difference in infant mortality between the two groups, we would be in a good position to propose a causal relationship. Of course, such an experiment is out of the question on ethical grounds and could never be done with humans. For this reason, we must be satisfied with relational rather than causal data on many issues of drug effects in humans.

## THE STUDY OF BEHAVIOR

### Behaviorism

Contemporary psychology should not really be called "psychology" at all. The name is derived from the Greek word *psyche*, meaning "mind," and it literally means "the study of the mind," which it is not. Modern psychology has its roots in both physiology and philosophy. Its name is derived from philosophy, but its methods and subject matter are closer to physiology.

Early psychologists attempted to study the mind scientifically, but they met with very little success because they used introspection as their method of gathering data. Introspection is the internal observation of what is going on in one's own mind. Introspection failed as a scientific method because it violated one of the principal requirements of science: The subject matter must be public—the data must be available for everyone to observe. Introspection, by its very nature, could not be a source of scientific information.

In the early part of the twentieth century, an American psychologist, John B. Watson, founded a school of psychology known as *behaviorism*, which has become the basis of contemporary Western psychology. Watson rebelled against any study of the mind or of any other concept that could not be defined in terms of observable and measurable events. He claimed that the real subject matter for psychologists should be behavior. Behavior is public, measurable, and reproducible and, consequently, is amenable to the scientific method. As a result of Watson's influence, most experimental psychologists today study overt behavior, although the old terms *psychology* and *psychologist* are still used as widely as ever.

Psychologists and pharmacologists who study the effects of drugs on behavior sometimes refer to their field as *psychopharmacology*, which is simply a combination of the words *psychology* and *pharmacology*. Purists prefer to use the term *behavioral pharmacology*. Today the term *psychopharmacology* is most often used to describe the field that studies the effects of drugs on psychiatric symptoms—a merging of the concerns of both psychiatry and pharmacology.

### Introspection

***Unstructured Introspection.*** Introspection, the early psychologists' method of studying internal mental processes, did not succeed as a scientific method and is not particularly helpful as a tool in behavioral pharmacology. This is not to say that unstructured verbal descriptions of drug-produced internal states are not useful to a researcher. On the contrary, they guide and inspire more systematic study. But the accounts themselves are not adequate scientific data unless they are collected in a systematic or structured fashion.

Many chapters in this book provide accounts of the experiences of people who have taken various drugs. These can give us some indication of how a drug affected the writer at that time and can be fascinating, but by themselves they can give readers some indication only of how they might feel if they took a similar dose of the drug. These accounts alone cannot provide data that can be used to predict or understand the action of the drug in terms that scientists can find useful.

***Systematic Introspection.*** Introspection by itself is of no value to the behaviorist, but the acts of introspecting and the subsequent verbalization

are behaviors that can be studied. If a psychologist sticks a pin into a subject's finger, the pain will be quite real to the subject but is available only to the subject—not the experimenter—for study. However, if the psychologist asks, "Does it hurt?" and the subject answers, "Yes," the answer is public and scientifically useful. Similarly, when a drug is given to an experimental subject who hallucinates, the visions seen are the subject's own and cannot be studied, but the subject's responses to the visions can be recorded, analyzed, measured, and subjected to any kind of behavioral test. The scientific usefulness of such responses will, to a large extent, depend on the care with which they are collected and the skill with which they are analyzed.

Systematic self-reporting is also useful in human research when studying the effects of drugs on internal states such as moods. Rather than just asking subjects how they feel, the researcher can use scales that measure a particular aspect of how the drug makes a person feel. One example of such a scale is the Profile of Mood States (POMS), a paper-and-pencil test that asks subjects to describe, using a five-point scale, how each of 72 adjectives applies to them at a particular moment. These 72 items yield a score on eight independent subscales: anxiety, depression, anger, vigor, fatigue, confusion, friendliness, and elation. These scales give a reliable and quantifiable measure of a subject's internal state (see Chapter 11 for an example).

The Addiction Research Center Inventory (ARCI) is a similar test developed specifically to assess the abuse potential of drugs. The complete questionnaire consists of 550 true–false items covering a broad range of physical and subjective effects. This scale can be given to a person who has taken a new drug, and the drug's subjective effect can be compared with the effects of these other drug classes (Haertzen & Hickey, 1987). The assumption is that drugs with similar subjective effects, particularly euphoric effects, will be similarly abused (Preston, Walsh, & Sannerud, 1997, pp. 100–103).

## LEVEL OF AROUSAL

The concept of *arousal level* has been around for a long time and is the basis for a number of important psychological theories. In the course of a day, we all experience changes in arousal that normally range from deep sleep to mild excitement. Arousal level is thought of as one continuous variable so that the arousal level at a specific time can be characterized as being at a point on a scale and moving in either direction. Figure 2-1 shows the scale of arousal level and indicates the normal range of arousal we encounter in the course of a day.

Arousal can continue above or below the normal ranges, causing pathological or abnormal states. At the low end, arousal goes beyond deep sleep to coma and death; at the high end, excitement ranges into mania, convulsions, and death.

One of the oldest ways of classifying drugs is by whether they increase or decrease arousal. Drugs that increase arousal are "stimulants" or "uppers," and those that decrease arousal are "depressants" or "downers." This distinction, however, is of limited usefulness in behavioral pharmacology because it really describes only the effect of a drug at high doses. Depressants at high doses cause loss of consciousness, and stimulants at high doses cause convulsions. At intermediate doses that are of interest to the behavioral pharmacologist, different stimulants have quite different

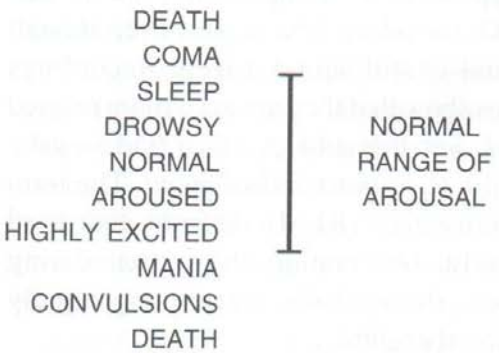


**Figure 2-1** The range of arousal levels. Normal arousal varies between deep sleep and excitement, but drugs and disease may cause arousal to move out of the normal range.

effects, and the same is true for depressants. In fact, most "depressant" drugs stimulate spontaneous motor activity at low doses. In fact, it is a principle of behavioral pharmacology that the effect that any drug has on behavior depends far more on the behavior being measured and on the organism and its behavioral history than on the nature and dose of the drug.

The *electroencephalograph* (EEG) is a device that measures electrical waves produced in various parts of the brain. It is described in more detail in Chapter 4 in the section on brain imaging techniques. Apart from detecting activity in different parts of the cortex, it is often used as a measure of arousal. In an alert individual, the EEG will usually show *beta waves*, and a relaxed individual with closed eyes will show smooth, regular waves called *alpha waves*. As a person becomes more relaxed, the waves get larger and the cycles longer.

### Sleep

During sleep, these waves become very slow and have a very large amplitude and are called *delta waves* or sleep spindles. As a person goes to sleep, five distinct stages can be recognized with the EEG. These stages are characterized by slower and slower frequency, except the last stage, known as paradoxical or REM sleep. Over the course of a night, the EEG reading shows the brain cycling through these five stages. Every 90 minutes or so, the brain appears to wake up; it shows a normal waking EEG, including beta waves, even though the individual is still sound asleep. Recordings from muscles show that they are even more relaxed than normal, yet the subject has a wide-awake EEG, thus the term *paradoxical sleep*. The term *rapid-eye movement* (REM) sleep is also used because careful observation shows that during these periods, the eyeballs are moving rapidly under the closed eyelids.

The physiological and psychological significance of REM sleep is not understood, but it is known that dreaming is associated with REM, and people deprived of REM sleep suffer from changes in intellectual ability, motivation, and personality. REM sleep can be altered by various drugs.

### Arousal and Mood

Drug-produced arousal-level changes are often mistakenly associated with moods. The colloquial classification of drugs as "uppers" and "downers" is based on this presumed relationship between mood and state of arousal; that is, an upper makes you feel up or good, and a downer may not be associated with pleasure or happiness. However, it cannot always be assumed that arousal has anything to do with mood. For example, the stimulation and jitters caused by many stimulant drugs such as caffeine are considered unpleasant, as is the depressed, sedated state caused by the antipsychotic drugs or the confusion caused by benzodiazepines. By contrast, opiates, which cause a depressed, dreamy state, and cocaine, which causes high arousal, produce euphoria and extremely pleasurable mood states called *rushes*.

## MEASURING PERFORMANCE IN HUMANS

Some tasks, such as driving a car, are of great interest to behavioral pharmacologists and are studied through elaborate tests and simulations of real-life conditions. Although findings from these tests are directly applicable to the task studied, it is generally better to study the basic intellectual and behavioral processes used in these more complex skills and apply the findings more generally to a greater variety of tasks that use the same basic skills. Most tests of human performance can be categorized according to whether they measure perceptual, cognitive, or motor abilities.

***Perception.*** A number of tests and techniques have been developed to measure the acuity of the senses, particularly sight and hearing. Sensitivity changes are reported as changes in thresholds. The term *absolute threshold* refers to the lowest value of a stimulus that can be detected by a sense organ. It is a measure of the absolute sensitivity

of the sense organ. *Difference thresholds* are measures of the ability of a sense organ to detect a change in level or locus of stimulation. If a threshold increases, it means that the intensity of the stimulus must be increased in order for it to be detected. In other words, the sense has become less keen. A lowering in threshold means that a sense has become more sensitive.

An example of how threshold is measured is critical frequency at fusion. If the speed with which a light flickers is increased, eventually a point will be reached where the light appears to be steady. This is the critical frequency at fusion, and it is sensitive to many drugs. The ability to detect flicker is a reliable measure of how well the visual system is functioning. To measure the functioning of hearing, an auditory flicker fusion test has also been developed.

***Cognitive Performance.*** Some very sophisticated tasks have been developed over the years to assess cognitive abilities, that is, the ability to process, store, and retrieve information. Tests of vigilance or attention assess the ability to detect the presence of a specific type of signal in a group of signals, often over an extended period of time. Tests of memory and of learning and recall require a subject to learn a list of items that must be recalled at a later time.

The entire process of learning and testing may take place while the subject is under the influence of a drug, or the effect of a drug on different stages of the process can be determined by giving the drug during either the learning phase or the recall phase.

Other tests of intellectual functioning include the ability to perform addition and subtraction mentally, to sort cards, and to name colors. A commonly used paper-and-pencil task is the *digit symbol substitution test* (DSST). It was originally part of a commonly used intelligence test. At the top of the page in the DSST, the subject is given a table that shows a symbol associated with nine digits. The task is to draw the appropriate symbol next to each number in an array of numbers as fast as possible. It tests memory and psychomotor skills.

***Motor Performance.*** Motor performance is a major concern in assessing the effects of drugs on human performance. Some of the simplest measures of performance are reaction time, tapping rate, hand steadiness, and body sway. Hand–eye coordination is often measured by a device called a *pursuit rotor*. With this device, the subject is instructed to hold the end of a stylus on a point on a rotating disk. The total time the subject is able to hold the stylus on the moving spot is a measure of hand–eye coordination.

***Driving.*** Because driving is such a necessary and common activity, it is important to know the effect many drugs have on the ability of a person to operate an automobile. Determining the effect of a drug on driving ability, however, is not as easy as it might seem. To begin with, driving is a complex activity requiring many skills of perception, motor control, and judgment. There is much more involved than simply moving a car from one place to another. Researchers have tried to assess driving skill by using many different strategies. Some simply have their subjects drive a car through city traffic and have professional driving instructors rate their performance on a number of factors. One difficulty with this approach is that the demands of the task will be different for each person tested because traffic conditions are constantly changing. However, the main problem is that it is unethical to permit subjects to drive in real traffic and endanger their lives and the lives of others if there is any possibility that their skills might be impaired by drugs.

To get around these problems, researchers sometimes have their subjects operate a vehicle around a closed course where various demands are made on the skill of the driver. This approach is more artificial but safer, and because the task is the same for each subject, comparisons are more easily made between and within subjects.

One difficulty with using a real car is that it is sometimes difficult to measure a subject's performance accurately. You can tell if the subject knocks over a pylon, but you will not be able to determine whether the error resulted because the object was

not seen, the subject could not estimate the speed of the car, or the reaction time was too slow. To answer such questions, many researchers use computerized driving simulators that are capable of measuring a subject's response time, steering ability, and capacity to react to specific crises. With some simulators, it is even possible to measure the subject's eye movements while driving.

## MEASURING BEHAVIOR IN NONHUMANS

### Unconditioned Behavior

The simplest measure of behavior in nonhumans is how much of it there is. Such measures are usually referred to as *spontaneous motor activity* (SMA), which may be measured in a number of ways, but usually the animal is placed in an "open field" (a large open box), and its movements are measured either electronically or by drawing a grid on the floor of the open field and counting the number of times the animal crosses a line.

Much can also be learned simply by observing the behavior of animals after they have been given drugs. Some classes of drugs cause animals to exhibit stereotyped behavior—the continuous repetition of a simple act such as rearing or head bobbing. Other drugs may cause sleep or convulsions.

It is also possible to measure other unconditioned behavior using very simple techniques. For example, muscle tone in rats can be measured using an *inclined plane test* where the animal is placed on a board that can be tilted to various degrees. The degree of tilt where the animal is unable to hold on to the surface and slides off is a measure of muscle tone.

A test used to measure anxiety is the *elevated plus maze*, which consists of narrow boards in the shape of a "plus" raised a foot or so off the ground. Two opposite arms of the plus have walls, and the other two do not. Normally, rats spend most of their time on the arms that have walls and only occasionally venture out on the unprotected arms. Drugs that are known to relieve anxiety cause rats to spend more time on the unprotected arms than they normally would.

There are several tests for *analgesia*, or the ability of a drug to block pain. The most common is what is referred to as the *paw lick latency test*. Rats are placed on a metal surface that is heated to about 50 degrees Celsius. This is about the temperature of a hot cup of coffee. When you first pick it up, it feels warm, but the longer you hold it, the hotter it becomes, until you have to put it down. When rats are first placed on the hot surface, they do not react, but within a few seconds, they raise one of their hind paws to their mouth as though they were licking it. The length of time it takes for this to happen is called the *paw lick latency*. Analgesic drugs like morphine lengthen this latency, which is often used as a measure of a drug's analgesic effect. Even if the rat does not show this response, it is removed after a fixed number of seconds to prevent the heat from burning the skin.

### Conditioned Behavior

Learned behavior is frequently classified by whether it is a result of *respondent* or *operant* conditioning. This distinction does not represent two types of learning; rather, it arises from attempts to condition two different types of behavior, reflexive and voluntary. Respondent conditioning is also known as *classical* or *Pavlovian* conditioning because it was the first type of learning to be studied systematically and was first investigated by Ivan Pavlov, the great Russian physiologist.

### Respondent Conditioning

Respondent behavior is reflexive in the sense that it is under the control of well-defined stimuli in the environment. When a dog salivates at the sight of food, the salivation is respondent behavior under the control of the stimulus of food; that is, it is a reflex. Such a reflex is considered to be unconditioned. Thus, the stimulus, the food, is the *unconditioned stimulus* (UCS), and the salivation is the *unconditioned response* (UCR). Pavlov found that if the sight of food is paired with a neutral stimulus

such as a ringing bell, the bell alone eventually elicits the salivation. Thus, the bell becomes a *conditioned stimulus* (CS), and the salivation to the bell in the absence of the food is the *conditioned response* (CR). The process by which stimuli become conditioned has been extensively studied, and a great deal is known about the conditions that control the development and expression of respondent conditioning. Pavlov's studies of drugs' effects on respondent conditioning were some of the very first experiments in behavioral pharmacology. Such studies are rare in behavioral pharmacology today. The importance of respondent conditioning is that it can help us understand many of the effects that drugs have, as we shall see in Chapter 3.

### Operant Conditioning

*Operant behavior* is behavior that appears to be voluntary because it is not elicited by any apparent stimulus in the environment. Operant behavior may be conditioned if it is followed by a *reinforcing stimulus*, such as food. A dog that learns to beg for food at the table is demonstrating operant conditioning. The begging is the operant, and it is maintained by the occasional delivery of food. If begging no longer results in the delivery of food, the begging stops.

The principles of operant conditioning are thought to apply to nearly all behavior of all animals. Operant behavior is usually studied in the laboratory using a *Skinner box*—a small cage attached to an apparatus that will deliver small quantities of food or water. It also contains a *manipulandum*, something that the animal can manipulate or move (e.g., a bar, lever, or knob). Figure 2-2 shows a Skinner box for a rat. In this box, a food delivery system delivers one small pellet of rat food at a time. The manipulandum is a lever on the wall near the food dish.

To study operant conditioning, it is usual to first deprive the subject of food or water so that these can act as rewards for performing the desired operant (in this case, pressing the lever). Each lever press is detected electronically and causes food to be delivered. In this way, the rat is rewarded, or *reinforced*, each time it makes the desired response. When the rat has learned this response, it makes it frequently and reliably. However, the first time it is placed in the box, it may not depress the lever on its own, so a procedure called *shaping* is used. In shaping, the researcher first reinforces the rat for going near the lever. Then, when the rat is approaching the lever reliably, the researcher waits until the rat touches the lever before delivering food. Eventually, the behavior of the rat is shaped by rewarding closer and closer approximations of the desired behavior until the rat is pressing the lever and being rewarded electronically.

**Figure 2-2** A Skinner box for a rat. (Courtesy of Gerbrands Corporation, Arlington, Massachusetts 02174)

## Schedules of Reinforcement

To maintain this behavior, it is not necessary to reward the rat with food every time it presses the lever. Animals will usually respond many times for one reinforcement, and in most operant research, this is the case. Reinforcement may be given after a specific number of responses or on the basis of time. The term *schedule of reinforcement* refers to the pattern that determines when reinforcements are to be given.

Each schedule of reinforcement engenders a characteristic pattern of responding that will be seen no matter what the species or organism or the type of reinforcer. These patterns are reliable and predictable, and they are sensitive to the effects of many drugs. Behavioral pharmacologists have found them a useful means of analyzing the behavioral effects of drugs because specific schedules are more sensitive to some drugs than others, and similar drugs affect schedule-controlled behavior in a similar manner.

Not only do schedules provide a powerful method for analyzing and classifying drugs, but the study of the effects of drugs on operant behavior is useful in other respects. It is believed that most human behavior, although very complex, is ultimately controlled by reinforcements, just as the behavior of the animal in a Skinner box is. It is believed that by carefully studying the effects of a drug on operant behavior, we can provide valuable information that will help us understand the effects of the drug on the behavior of humans.

***Ratio Schedules.*** When reinforcement is based on the number of responses an animal makes, the schedule is known as a *ratio schedule*. On a *fixed ratio* (FR) schedule, the animal is required to make a fixed number of responses in order to be reinforced. For example, on an FR 30 schedule, every 30th response produces a reinforcement. If only 29 responses are made, the reinforcement is never given. A *variable ratio* (VR) schedule is similar except that the number of required responses varies randomly so that at any given time, the occurrence of a reinforced response cannot be predicted. A VR 30 schedule will produce a reinforcement every 30 responses, on the average.

***Interval Schedules.*** On an *interval schedule*, an animal's responding is reinforced only if a period of time has elapsed since some event, such as a previous reinforcement. Responses that the animal makes during that time are recorded but do not influence the delivery of reinforcement. On a *fixed interval* (FI) schedule, a response is reinforced only after a fixed time has elapsed. A typical example might be an FI 3 schedule; the animal must wait 3 minutes after the delivery of a reinforcement for a response to be reinforced again. On a *variable interval* (VI) schedule, the interval is randomly determined. When a value is specified for a VI, such as VI 2, this indicates that the interval is an average of 2 minutes long.

***Avoidance–Escape Task.*** Not only will animals learn to press a lever to obtain appetitive rewards like food, but they will also learn to avoid and escape from unpleasant stimuli such as electric shocks. On a typical avoidance–escape schedule, the animal is presented with a stimulus such as a buzzer or a light as a warning that a shock is coming. The warning comes several seconds before the shock, and if the animal makes a response during that time, the warning stimulus is turned off, and the shock never comes. If the animal does not respond, the shock comes on, and the animal can then escape from the shock by responding.

This type of task has proved to be a valuable tool in analyzing the effects of drugs. It has been found that drugs useful in treating anxiety in humans interfere with an animal's ability to avoid shock. These drugs block avoidance behavior during the signal but do not have any effect on the animal's ability to turn off or escape from the shock when it does come. This finding shows that the drug has not interfered with the ability of the animal to respond but has selectively blocked the motivation to avoid the shock.

***Punishment.*** Responding that is maintained by appetitive rewards may be suppressed if it is also punished with electric shocks. This behavior is

usually measured by having the animal respond on a VI schedule that produces a steady rate of responding. At various times during a session, a stimulus is turned on and lasts for a minute or two. During this stimulus, each response is followed by a shock. Responding during this stimulus will be suppressed. The frequency and intensity of the shock can be manipulated to produce a specific amount of suppression. Some varieties of drugs, such as barbiturates, increase the frequency of behavior that has been suppressed by punishment. Other drugs, such as amphetamine, lack this ability.

As with classical conditioning, a great deal of research has studied the effects of drugs on operant conditioning and schedules, but operant conditioning principles also play a role in understanding the effects of drugs—in particular, why and how organisms use drugs (see Chapter 5).

## DISSOCIATION

Material learned while under the influence of a drug may not be easily recalled after the effects of the drug have worn off, and, conversely, information learned while drug free is not easily recalled when intoxicated. This phenomenon, known as *dissociation*, although recognized many years earlier, has been systematically investigated only since 1945.

The first account of dissociation appeared in 1830 in a British medical text. It was related by George Coombe, who stated that "before memory can exist, the organs require to be affected in the same manner, or be in a state analogous to that in which they were, when the impression was first received" (quoted in S. Siegel, 1982, pp. 257–258). As an example of this phenomenon, Coombe cited the case of an Irish porter who had lost track of a valuable parcel while drunk. Later, while sober, he was unable to remember anything about it. The next time he was intoxicated, however, he remembered exactly where he had put it.

Dissociation research often uses the following design. Experimental subjects are assigned to four groups, and they all learn a task on day 1 and are tested for recall on day 2. Group 1 gets a drug on both training and testing days, group 2 gets the drug during training but not during testing, group 3 gets no drug on training days and the drug during testing, and group 4 does not get the drug on either day. The design and results of this type of experiment are shown in Table 2-1. Typically, subjects in groups 1 and 4 on day 2 are able to recall the training they received on day 1, but the subjects in groups 2 and 3 have difficulty remembering what they learned on day 1 because of the change in drug state between these 2 days.

Dissociation has been demonstrated in a number of species, with many drugs, and using a variety of memory tests (Overton, 1972), and it may actually be a cause of much day-to-day forgetting. It has been shown that many drugs we consume regularly every day in moderate doses (e.g., in coffee, tobacco, and alcohol), especially when consumed together, can cause dissociation in a memory task that involved learning a route through a simple map (Lowe, 1988).

Dissociation may play a role in interpreting results of behavioral pharmacology experiments. Performance on a task may deteriorate when drugs are given or taken away. Such effects may be due not to an effect of the drug on the task but rather to a general effect of state change, that is, an inability to transfer learned information from one state (drugged) to another state (undrugged) or vice versa.

**TABLE 2-1 The Design of a Typical Dissociation Experiment**

On the training day, subjects are taught a task, and on the testing day, they are tested for recall of that task. For groups 1 and 4 the state does not change, but subjects in groups 2 and 3 are tested under drug states different from training. If the drug causes dissociation, subjects in groups 2 and 3 have difficulty remembering.

| *Group* | *Training Day* | *Testing Day* | *Result* |
|---|---|---|---|
| 1 | Drug | Drug | Memory |
| 2 | Drug | Saline | No memory |
| 3 | Saline | Drug | No memory |
| 4 | Saline | Saline | Memory |

It has been speculated that dissociation also may have an important influence on drug use because it means that much of what we learn about drug taking while we are not under the influence of the drug may not be easily remembered after we take the drug. An alcoholic, for example, may experience the adverse and punishing consequences of drinking while sober but may not remember these consequences while drunk. Drug education programs are always aimed at people while they are sober. Information acquired in this state may have little influence on a person after he or she takes a drug.

## STIMULUS PROPERTIES OF DRUGS

Another method of analyzing the effects of drugs is by their *stimulus properties*, their ability to act as a discriminative stimulus in a discrimination learning task. In this type of task, a hungry animal, usually a rat, is given a choice of two levers to press. In some sessions lever A will be reinforced, and in other sessions lever B will be reinforced. The reinforcement is on an FR 20 schedule so that without a cue to guide it, the rat will not know which lever will produce the food until it has made at least 20 responses. In this situation, the only cue is the presence of a drug. On days when lever A is reinforced, the rat is injected with a drug, but on days when lever B is reinforced, it is injected with saline. After a short period of training, the rat will learn to discriminate between the drug and saline. On drug days, it will start off responding on lever A, and on saline days, it will start off on lever B. Thus, the first 20 responses on any given day will show whether the rat thinks it has been injected with the drug or saline.

Animals are capable of discriminating even low doses of most drugs, but this is not the most interesting aspect of drug state discrimination. We can do several things in this experiment that will tell us a great deal about the drug. One often-used test is to give the rat a different drug and see how it causes the rat to behave. If the rat makes the response usually made to the training drug, it is said to generalize to that drug, and this result tells us that the two drugs are perceived as similar by the rat. Usually, rats will generalize to drugs that are members of the same class. These responses can be helpful in screening new drugs.

Drug state discrimination studies are also useful in determining the biochemical mechanisms by which a drug produces its subjective effects. This is done primarily through the use of antagonist drugs. For example, it can be demonstrated that the stimulus properties of amphetamine can be blocked by the drug AMPT, which blocks the enzyme that creates the catecholamine neurotransmitters (see Chapter 4). This finding tells us that the subjective effects of amphetamine depend on catecholamines in the brain.

Similar techniques have been used in human research. Subjects are given two capsules and asked if they can tell which one contained a drug similar to a previously administered test drug.

## DEVELOPMENT AND TESTING OF PSYCHOTHERAPEUTIC DRUGS

Before any new drug can be put on the market, it must undergo rigorous development and testing to prove that it is effective and safe. Only then will it be approved by governmental agencies for sale as a medicine. In the United States, approval is granted by the federal Food and Drug Administration (FDA). Other countries have similar agencies.

We do not understand the biochemical basis of mental illness well enough to specifically design drugs in the laboratory with any certainty that they will have a desired effect on psychiatric symptoms and will produce a minimum of side effects. Instead, the laboratories of pharmaceutical companies synthesize many new chemicals they think might be effective. These drugs are then screened using nonhumans to determine whether they have effects similar to those of known therapeutically useful drugs, and whether they are safe.

Screening tests, using laboratory animals, can determine whether a drug might have therapeutic properties (Leonard, 1989). It is known, for example, that drugs that are useful in treating anxiety or psychosis will block avoidance responding in doses that have no effect on escape responding. Drugs that are useful in treating depressive illness will block the depression in behavior caused by reserpine. In fact, many of the procedures described earlier in this chapter are used in the initial screening process of behaviorally active drugs.

As described in Chapter 1, toxicity of a drug is assessed by comparing the $LD_{50}$ and the $ED_{50}$ of the useful effect of the drug. The farther the lethal dose is from the effective dose—the higher the therapeutic index (see Chapter 1)—the safer the drug.

When a new drug appears to be reasonably safe and shows interesting behavioral properties in nonhumans, it goes to phase 1 of human testing, which assesses the toxicity and side effects of the drug on healthy human volunteers. In phase 2, the drug is tested on patients under very carefully supervised conditions. If phases 1 and 2 show that the drug has minimal toxic effects and also has a potential therapeutic effect, it then goes to phase 3, expanded clinical trials, which are usually carried out in university teaching hospitals and other institutions and often use the three-groups design discussed earlier. This research is usually carried out by a number of independent investigators under contract to pharmaceutical companies. In 2002 a problem with the reporting of clinical trials came to light. To test the significance of the difference between the treatment group and the placebo group, statistical tests are used. If the difference between the drug treatment group and the placebo group is so great that it would be expected less than 5 percent of the time, the treatment effect is considered meaningful. What this means is that if you repeat the trial 20 times, you would expect to see at least one significant result. If you publish that one trial without mentioning the 19 nonsignificant trials, it makes your drug look like it is effective when it is really not. To get around this problem, drug companies have agreed to register all clinical trials publicly before they are conducted and post the results even if they are negative so that everyone will know about them (see Box 2-2 for a description of how this came about).

If phase 3 investigations are successful, the drug is licensed and marketed. The research, however, does not stop here. Phase 4 involves the

**BOX 2-2** Registration of Clinical Trials

It has been estimated that only half of the clinical trials that have been conducted over the past 56 years have ever been reported, and many that were reported did not appear on a searchable index like MEDLINE, so they cannot be easily found (Rennie, 2004). Does this really make a difference? Is it really important that all clinical trials, even those that show negative results, be made public? It certainly is not in the interest of the drug manufacturers if they do an experiment and find that a new drug is no more effective than a placebo. It is also not in the interest of researchers and medical journals to use up valuable page space to publish negative or equivocal data since these findings are not likely to change clinical practice or be cited by other authors. As a result, the studies that do get published have a persistent bias toward positive results.

There are some unfortunate consequences of this tendency for only positive findings to be publicly available, the most serious of which is that harmful effects turned up by these studies are never made public. Here is a recent example.

(*Continued*)

**BOX 2-2**

In 2003 and 2004, there was a serious debate as to whether children and adolescents with major depression could be safe and effectively treated with selective serotonin reuptake inhibitors (SSRIs). There was a concern that SSRIs might actually increase the incidence of suicide attempts in this age-group, but no data were available. The drug company that manufactured paroxitine (Paxil), GalaxoSmithKline (GSK), had done two clinical trials on young people in the 1990s that were unable to show that paroxitine was more effective than a placebo. But they were unwilling to share all the results of their clinical trials with other researchers and appeared to have deliberately suppressed making them public for "commercial" reasons (Rennie, 2004).

On June 2, 2004, the attorney general of the state of New York sued GSK saying that the child clinical trials and safety outcome data from other trials had been hidden by GSK and that they had "deprived physicians of the information needed to evaluate the risks and benefits of prescribing paroxitine to children and adolescents with MDD [major depressive disorder]" (Rennie, 2004, p. 1360).

On August 26, 2004, New York Attorney General Eliot Spitzer announced a settlement with GSK where they agreed to putting summaries of the results of all clinical drug trials into a register posted on the Internet with a conspicuous link from their Web site. To see this registry, go to http://ctr.gsk.co.uk/welcome.asp.

In addition, an organization of the publishers of many prestigious medical journals, including the *New England Journal of Medicine* and the *Journal of the American Medical Association* (JAMA), have adopted a new policy that they will not publish any clinical trial that has not been registered in a public registry before it was begun (De Angelis et al., 2004). This is to ensure that drug companies will not "cherry-pick" the best studies for publication and be able to hide studies that do not turn out to be favorable to their product.

accumulation of data on the success of the drug as used in the clinic. Attempts are made to identify adverse effects that were not apparent in the short-term testing during the early stages. In phase 4, improved dosing schedules may be developed, and individuals who are at risk of having adverse reactions to the drug can be identified (Baldessarini, 1985, p. 5).

## CHAPTER SUMMARY

- Scientific experiments consist of an independent variable that is manipulated by a researcher and a dependent variable that is measured. In most experimental research in behavioral pharmacology, the independent variable is the presence of a drug in the body, and the dependent variable is some aspect of behavior.
- Treatment of control subjects in an experiment should be as similar as possible to treatment of experimental subjects. For this reason, control subjects are usually given a placebo, an inactive substance administered in exactly the same way as the drug. This procedure controls for differences that result from the placebo effect.
- The placebo effect refers to the observation that when people expect a drug, they often show the effect of a drug even if they are administered only an ineffective substance.
- Experimental and control conditions may be administered to the same subjects on different occasions (within-subjects design) or to different subjects (between-subjects design).

- To determine whether the differences between experiment and control conditions mean anything, a statistical test is often used. Statistical tests tell the researcher how frequently there would be differences as large as in the experiment if the drug had no effect at all and all differences were due to chance alone.
- A double-blind procedure, where neither the researcher nor the experimental participant knows which group they are in, is used to eliminate the effect of experimenter bias.
- The three-groups design is used in testing new therapeutic drugs. In this design there is a drug group, a placebo group, and a group that receives an established treatment.
- In nonexperimental drug research, when relationships between two measured variables are found, one cannot assume the existence of any causal relationships between variables.
- Although the name *psychology* literally means "the study of the mind," modern psychology is the study of behavior. Introspection may be used as a source of data about drugs, but such data must be collected systematically and interpreted carefully.
- Level of arousal ranges from deep sleep to mild excitement but may extend into pathological states of mania and convulsions at the upper end and coma and death at the low end. Arousal level cannot be assumed to be related to either mood or general activity.
- The performance of the senses is determined by measuring their threshold. Cognitive performance is measured by tests of the ability to process, store, and retrieve information. Motor performance may be measured by simple or choice reaction time, tapping rate, hand steadiness, or pursuit rotor tests.
- Effects of drugs on unconditioned behavior of nonhumans can be measured by observing the amount of activity in an open field or using an inclined plane or an elevated plus maze. Analgesic effects can be measured as paw lick latency on a hot plate.
- Conditioned behavior may be respondent or operant. In respondent conditioning, involuntary reflexive behavior is brought under the control of a previously neutral stimulus. This is also known as classical, or Pavlovian, conditioning. In operant conditioning, voluntary behavior is brought under control by delivery of a contingent reinforcement.
- When reinforcement is not given for every response but is given according to some pattern, the pattern is called a schedule of reinforcement.
- Animals can be trained to avoid and escape a noxious stimulus such as an electric shock.
- Dissociation refers to the fact that information acquired in one drug state may not be readily available in another drug state.
- Animals can be trained to make one response after being given a drug and a different response after being given saline or a different drug. The responses of an animal trained to discriminate a drug can be a useful tool in testing the biochemical mechanisms responsible for the subjective effects of a drug.
- Therapeutic drugs are first screened for effect and safety using nonhumans. They then go through four phases of testing on humans.