Internal Validity

If a study has internal validity, then you can be confident that the dependent variable was caused by the independent variable and not some other variable. Thus, a study with high internal validity permits conclusions of **cause and effect**. The major threat to internal validity is a **confounding variable**: a variable other than the independent variable that 1) co-varies with the independent variable and 2) may be an alternative cause of the dependent variable. For example, if you want to investigate whether smoking cigarettes causes cancer, you will need to somehow control for other possible causes of cancer that tend to vary with smoking. Each of these other possible causes is a potential confounding variable. If people who smoke more also tend to experience greater stress, and stress is known to contribute to cancer, then stress could be a confounding variable. Not every variable that tends to co-vary with the independent variable is an equally plausible confounding variable. People who smoke may also strike more matches, but if striking matches is not plausibly linked to cancer in any way, then it is generally not worth considering as a serious threat to internal validity.

Random assignment. The most common procedure to reduce the risk of confounding variables (and thus increase internal validity) is **random assignment** of participants to levels of the independent variable. This means that each participant in a study has an equal probability of being assigned to any of the levels of the independent variable. For the smoking study, random assignment would require some participants to be randomly assigned to smoke more than others. As this example makes clear, random assignment is not always ethical: if you are fairly confident that smoking does cause cancer, you could not ethically assign someone to smoke more. The usual alternative to random assignment is an **ex post facto** design, also known as a **classificatory** design, in which participants are assigned to levels of the independent variable not randomly but rather according to some characteristic that they already possess. For example, people who already smoke could be assigned to the smoking condition, and people who don't smoke could be assigned to the non-smoking condition. The problem with ex post facto designs is that these two groups may differ in many ways other than smoking. The smoking group may have higher levels of stress, they may have poorer quality health insurance, they may exercise less, etc. Any of these differences is a potential confounding variable. By randomly assigning participants to levels of the independent variable, you *reduce the risk of any systematic differences between the groups* before the study begins. Two groups that have been randomly assigned should, on average, be approximately equal to one another on most potential confounding variables. Because the only systematic difference between the groups is the one the experimenter controls – the independent variable – any differences between the groups in their dependent variables can be attributed to the independent variable. Random assignment is the sine qua non, or defining feature, of **experiments**. If a study does not use random assignment, it cannot be called an experiment. To review:

1. Internal validity is the confidence you can have that the independent variable is responsible (caused) changes in the dependent variable.
2. Random assignment increases internal validity by reducing the risk of systematic pre-existing differences between the levels of the independent variable.
3. Studies that use random assignment are called experiments.

In addition to the unique confounding variables that may threaten particular studies (e.g., stress threatening research on the smoking-cancer link), there are several threats to internal validity that are more general:

Participant expectancy effects. Participants in your study may discover what your study is about and act differently. The features of a study that reveal the hypothesis are called **demand characteristics** because they are thought to "demand" a particular response from participants. For example, if participants learn that they will be in a study on conformity, they may deliberately conform less than they would have otherwise. If the participants in different levels of the independent variable have different expectations of how they should respond to the independent variable, then these expectations are a potential confounding variable: observed differences in the dependent variable among the groups may actually be due to expectations rather than the independent variable. An example of this is the placebo effect. A **placebo** is a "false treatment" that is designed to look like a real treatment. Drug researchers

often give one group of participants a pill with some inert substance in it and the other group the real drug. The pill with the inert substance is called a placebo. A **placebo effect** occurs when merely expecting a certain effect is sufficient to produce the effect. Placebo effects have been observed for intoxicating substances, painkillers, mood enhancing drugs, and even hallucinogens. Given the power of participant expectations, it is important that participant expectations be carefully controlled in an experiment. The simplest way of controlling expectations is with a **blind study**. In a blind study, participants are kept blind to the level of the independent variable to which they have been assigned. They may know that they are watching a videotape, for example, but they would not be aware they are watching the "high-status speaker" version of the videotape while other participants are watching the "low-status speaker" version.

Experimenter expectancy effects. Rosenthal and Fode (1963) conducted a study in which research assistants were told that they would be training rats for a study on maze performance. Some assistants were told that their rats were "maze-bright" while others were told their rats were "maze-dull." In fact, all of the rats were carefully screened to have the same ability – only the assistants' expectations differed. When the day came to test the rats, however, a change was observed: the "maze-bright" rats were faster than the "maze-dull" rats. How could this be, if the rats were all the same? The answer is that the research assistants treated the rats differently during the training period, unintentionally encouraging the maze-bright rats but not offering special assistance to the maze-dull rats. If rats respond to experimenter expectations, then the more cognitively sophisticated species of homo sapiens is likely to as well. Experimenters may inadvertently communicate demand characteristics to participants in a variety of ways – treating one group in a more tense or relaxed manner, taking one group more seriously than the other, etc. These differences in treatment could become confounding variables that threaten the internal validity of a study. To control for experimenter expectations, a **double blind study** is used. In a double blind study, neither participants nor the experimenters with whom they directly interact are informed about the level of the independent variable to which participants have been assigned. To create a double-blind study, researchers sometimes hire temporary research assistants who do not learn the hypotheses of the study until after the study is concluded. If it is not possible to use a double-blind study (e.g., because extra research assistants are not available), one way to reduce experimenter expectancy effects is to use a script to standardize interactions with participants.

Instrument decay or *instrumentation effects*. Instrument decay occurs when the standards of measurement change over time. An example might be a spring scale used at a loading dock. In the morning, the spring scale registers sacks of flour at approximately 50 pounds. By evening, the spring scale is registering the same sacks of flour at approximately 55 pounds. The standards of the spring scale have changed because the spring has become stretched out. Although the term *instrument decay* suggests mechanical instruments, it also applies to measurement by human judges. Judges may get better at measuring, they may get worse, or their standards may simply change because they attend to different aspects of performance (paying more attention to a diver's feet, for example). In every case, the measurements at different times may be due to changes in the measurer rather than actual changes in behavior, making them a potential confounding variable. If a researcher conducts all the control condition sessions first and all the experimental condition sessions last, then any differences between the control and experimental conditions may actually be due to instrument decay and not to level of the independent variable. For this reason, researchers should endeavor to balance or randomize the order in which experimental and control conditions are conducted.

Special threats to internal validity in within-subjects designs. There are many potential threats to internal validity with within-subjects designs, in which participants receive more than one level of the independent variable. These will be covered when we discuss within-subjects designs.

Alternatives to random assignment. There are many situations in which it is either unethical or impractical to randomly assign participants to levels of the independent variable. For example, it is not practical to randomly assign participants to income or place of residence, and it is not ethical to randomly assign participants to conditions that are dangerous (e.g., excessive alcohol consumption). When a researcher desires to maximize internal validity despite the unavailability of random assignment, **quasi-experimental designs** are often used. These designs attempt to reduce the risk of confounding variables through methods other than random assignment, and are discussed in the later chapter on quasi-experimental methods.