3

Designing Experiments and Conducting Simulations

The Experimental Method: An Overview

Although experimental methodologies have become very sophisticated during the latter part of the 20th century, only a few essential ideas characterize the approach. It has been regarded historically as being synonymous with the scientific method. Experimentation is a technology for assessing causality; it is supported by philosophical approaches that emphasize explanation. The *essentials* have not changed much since John Stuart Mill wrote about reconciling syllogistic logic with the methods of inductive science in his 1843 book titled A *System of Logic*. Favoring deduction over induction, experimentation has the following key features:

1. Experimental treatments can be applied independent of the prior states of the persons or groups being studied. This is accomplished through random assignment to the treatments being compared.
2. Experimental subjects are unaware of the purpose of the experiment or that other units are getting different treatments.
3. The experimental investigator has control over the administration of the treatments, referred to also as the *independent variables*. This administration occurs before measurements are made.
4. The same aspects of subjects' behavior, decisions, or perceptions are measured in each condition. These aspects, referred to as dependent variables, are free to vary.
5. Each condition of the experiment is repeated or replicated many times—usually at least 10 times—in order to assess variation in the dependent variables within and between the conditions.

These features facilitate the comparisons that are needed to infer causal relations between the treatments (independent variables) and the measurements (dependent variables). The key advantage of controlled comparisons is reduced ambiguity or increased confidence in specifying the direction of a relationship between variables. Confidence in making these inferences is increased to the extent that threats to internal validity are reduced (Campbell & Stanley, 1963). This subject is discussed widely in contemporaneous methods textbooks. I will have more to say about it shortly. But another advantage of experiments, discussed less frequently in texts, is the opportunity to make careful observations of the details of a process (such as negotiation) largely inaccessible in field settings. The human-subjects laboratory experiment provides a window into unfolding processes similar to the natural scientist's microscope. This advantage is discussed further in Chapter 9 on content analysis. The laboratory experiment is also a vehicle for exploring novel conditions and strategies in a relatively safe environment (Mahoney & Druckman, 1975, refer to this as a "test drive before you buy"). Many forms of third-party intervention in conflicts would benefit from a trial run in the laboratory before implementation in the field.

There is, however, another side to the well-known experimental coin, which is often used to assign subjects to conditions on a random basis. The setting is contrived: Experimental subjects are usually college student role players, time is often greatly compressed, past and future relationships are typically non-existent, issues and options in a conflict are usually assigned to subjects who play roles, and conflicts take the form of games or scenarios devised by the research team. These features pose a problem for generalizing results to real-world situations. Confidence in making the case for relevance turns on reducing threats to external validity (Campbell & Stanley, 1963). The issue is less about the obvious differences between the settings in scale and scope but whether these differences are critical variables in terms of influencing the experimental findings. A related issue is whether the experimental results suggest mechanisms that operate also in similar field settings. These issues can be understood in terms of trade-offs between internal and external validity.

Internal and external validity are important considerations in experimental design and analysis. They are discussed in the next section. The other major section of this chapter deals with the variety of designs and models that are relevant for addressing issues in CA&R. The discussion in these sections is intended to address questions often asked by new researchers in CA&R: Can experiments provide useful information about conflict and conflict resolution processes? Which designs are most appropriate for addressing particular issues in the field? What is involved in designing and implementing experiments? Understanding the strengths and limitations of experiments helps a researcher decide whether to use this approach to doing research. Knowing how to perform an experiment increases a researcher's confidence in using the approach. This chapter contributes to both these skills.

Internal and External Validity

The randomized control design gives an experimenter confidence in inferring causation from results of analyses, namely, that the manipulated treatment (problem-solving procedures) causes the measured outcomes (change in attitudes toward another group). This design has been regarded as a kind of panacea for the plausibility of alternative explanations for findings, namely, that another, extraneous factor that may be correlated with the treatment (contact with another group's members) caused the measured outcomes (changes in attitudes). But there is a limit to the number of controls that can be instituted in any experiment and, because of this, arguments that favor the hypothesized causal relationship between the treatment (problem-solving procedures) and outcomes (attitudes) can almost always be challenged. An experimenter's confidence in findings that support an hypothesized relationship between alternative treatments and outcomes increases to the extent that threats to internal validity are reduced.

The most well-known list of factors that may pose threats to internal validity was suggested by Cook and Campbell (1979). Among the 12 factors in their list are external events, referred to as *history;* aspects of measurement, referred to as *testing, instrumentation, and regression;* changes in the study's participants, referred to as *mortality or dropouts,* and *maturation;* and suspicions aroused by conversations among participants about the experiment, referred to as *diffusion of treatments, compensatory equalization of treatments,* and *compensatory rivalry* (see also Robson, 2002, for further discussion of these threats). I have added several other threats to the Cook and Campbell list. One of these is a participant's self-consciousness about being in an experiment, known as a *Hawthorne* or *placebo* effect. Another is suspicions about the experimental hypothesis aroused by subtle communications during the experiment, referred to as *demand characteristics.* A large research literature on the influence of expectations shows that experimenters often send signals, verbally and non-verbally, that influence subjects' behavior (see Harris & Rosenthal, 1985). Two other threats deal with measurement issues. One is the preference to choose socially desirable responses to questions about attitudes. Another is a preference for responding to items in a positive or affirmative direction. (For a discussion of these response biases, see Rokeach, 1963, and Druckman, 1970.)

Experimentalists have been inventive in developing controls for many of these threats, particularly those concerning dropouts, measurement issues, and response biases. Others have been more difficult to control, particularly those involving external events, maturation (in longitudinal experiments), and diffusion of treatments. In any event, it is not possible to perform an experiment in a vacuum; controls are a matter of degree. It is possible, however, to compare results obtained from experiments that control for different kinds of threats to validity. It is also possible to replicate experiments. Replicated findings—the same results obtained in repeated experiments—increase an investigator's confidence in them. Knowledge develops from programs of research, not from a few non-replicated experiments.

Confidence in experimental research findings is one part of the validity issue. The other part refers to relevance of the findings for other settings. Referred to as external validity, relevance may be limited by several features of an experimental design. LeCompte and Goetz (1982) suggest four threats to external validity. Findings may be limited to the particular groups studied (*selection*), the particular *setting* in which the study took place, or the specific *historical circumstances* that existed at the time of the study. These threats are magnified when college students play assigned roles in simulated conflicts. They are addressed, but not resolved, by assessments of similarity between the experimental setting and a comparable real-world situation. (A compelling case for external validity of an arms-control simulation was made by Bonham, 1971.) A fourth threat is that the particular *constructs* studied may be specific to the groups in the investigation. For example, problem-solving procedures may be understood differently by students in a role play and by policymakers wrestling with similar issues in actual conflicts. Two other threats to external validity are *investigator perspectives* and *non-representative sampling.* The former refers to the particular interpretive preferences held by the study's investigative team. This threat can be addressed by including different perspectives on the research team or by soliciting comments from colleagues on early drafts of articles. The latter refers to missing aspects of the population from which the experimental sample is drawn or is presumed to represent. This threat is reduced by choosing participants according to a sample design. This is accomplished to a degree when experiments are embedded in random-sample surveys (see Chapter 5 on survey research).

The two validities can be considered as being complementary. The more controls instituted in an experimental design to reduce threats to internal validity, the less the experimental setting is likely to resemble other (non-experimental) situations. Internal validity is often gained at the expense of external validity. In their attempt to reduce the plausibility of alternative explanations for findings, or to reduce threats to internal validity, experimentalists often simplify the tasks and settings of their experiments. A CA&R example is a simulation of the conflict between the Greek and Turkish communities in Cyprus (Druckman et al., 1988; Druckman & Broome, 1991).

In order to implement the Cyprus simulation in a reasonable amount of time (2 hours), a limited amount of background information about the conflict was presented, preparation time for the negotiations was limited, the role-playing students negotiated only three issues presented in the form of scaled options, and a time limit of 35 minutes was imposed on the negotiations. In order to control for the effects of testing, no attitude pre-test was administered; this was a randomized post-test design. Further, in order to separate understanding differences between the parties in values (the research question of interest) from familiarity or contact with the opponent, a simulated pre-negotiation workshop was described for one of the experimental conditions. An actual workshop would have required interaction before the negotiation.

These design features lend confidence in the findings at the expense of generality. The situation being simulated consisted of knowledgeable representatives of the parties, extended negotiations over many interrelated issues, changing attitudes over time, and participation in actual workshops in which both familiarity and liking are influenced. The controls instituted to reduce threats to internal validity (e.g., external events, testing, maturation, diffusion of treatments, and compensatory equalization of treatments) increased the threats to external validity (selection, setting, and history or a specific period of time). A more realistic simulation, on the other hand, would have raised questions of extraneous variables and comparability among the conditions. In this example, increases in one type of validity lead to decreases in another. This idea suggests that designers aspire to a balance between the validities. Rather than attempt to optimize either validity, an experimentalist can tweak both in order to achieve reasonable verisimilitude or generalizability without forfeiting interpretability of the findings. Some examples of designs that address issues of balancing are discussed next.

Experimental Designs: Examples From CA&R

Many CA&R researchers prefer the case study method. A question of interest is, why not perform experiments that address issues about conflict and conflict resolution? There are plenty of hypotheses to evaluate, and the method is relatively inexpensive and easy to implement. The methodology has the advantage of encouraging careful, precise thinking, usually in a deductive form. Randomized control group designs address issues of internal validity. Although threats to external validity are difficult to reduce, the issues are addressed with experiments that take the form of simulations: Experimental simulations are attempts to reproduce complex real-world situations and allow for a comparison with findings obtained in those settings. In this section, I introduce the CA&R researcher to alternative experimental designs and then discuss some simulations that have been used to evaluate hypotheses about conflict behavior.

A clear exposition of alternative designs can be found in Creswell (1994). His distinction between pre-experimental, quasi-, and classical experimental designs can be regarded as a progression from weak to strong attempts to reduce threats to internal validity. The pre-experimental design is particularly weak on comparisons or controls. It may take the form of a one-shot case study in which one treatment (designated as X) is evaluated (designated as O) with (O1 ——— X ——— O2) or without a pre-test (X ——— O). Or, it may compare non-equivalent groups where participants are not assigned randomly to a treatment or no-treatment group (Group A: X ——— O; Group B: ——— O), or two treatments are compared without random assignment (Group A: X1 ——— O; Group B: X2 ——— O).

Quasi-experimental designs add the feature of control groups. They may take the form of a pre-test, post-test comparison of two groups, one receiving a treatment, the other not receiving the treatment as follows:

Group A: O ——— X ——— O

Group B: O ——————— O

Because participants are not assigned randomly to these groups, the comparison is between non-equivalent groups. This feature enhances the plausibility of alternative interpretations for any differences found between the groups. Each alternative interpretation is a threat to the internal validity of the experiment. This sort of design is frequently used in field experiments where opportunities for random experiment are limited. For example, questionnaires that ask about attitudes toward the other group can be administered prior to (pre-test) and after (post-test) a problem-solving workshop (Group A). The same questionnaire can be administered, with the same time interval between administrations, to another group (B) not exposed to the workshop. It would be unlikely that participants are randomly assigned to these groups: Most workshops are intended to address relational issues between members of actual groups in conflict; random assignment to treatment and control groups would raise ethical issues in this context. It is also unlikely that Group B's activities between the pre- and post-test are similar to Group A's activities in all ways save for the problem-solving procedures. Thus, differences found in the change scores (pre- to post-test) can also be attributed to factors other than the workshop procedures.

A somewhat stronger quasi-experiment is the time-series design. Its strength derives from the feature of multiple measures before and after an intervention. Although randomization is missing, the time-series design enables a researcher to evaluate change in the context of a longer history of events or relations between groups. It also allows for an evaluation of the extent to which the change is sustained through time as well as compared to other groups repeatedly assessed without the intervention. Examples of time-series designs are discussed later in this chapter. (For more on a variety of quasi-experimental designs used in other areas of social science, see Cook & Campbell, 1979.)

The strongest design is the classical experiment. Random assignment (R) is intended to ensure that the comparison groups are equivalent. A popular form of the randomized control group experiment is called the Solomon four-group design. Participants are assigned randomly to one of four groups as follows:

Group A: R ——— O ——— X ——— O

Group B: R ——— O ————————— O

Group C: R ————————— X ——— O

Group D: R ————————————————— O

In this design, randomization ensures equivalent groups, and the groups without pre-tests (C and D) provide controls for the effects of testing. Supporting evidence for the hypothesis that the treatment (X) makes a difference would consist of the following results of group comparisons: Group A change scores (pre-test to post-test) are greater than (>) Group B change scores, Group C post-test scores are greater than (>) Group D post-test scores, and Group A post-test scores equal Group C post-test scores. These results would provide compelling evidence for the hypothesis, particularly because they eliminate an interpretation based on experience with a pre-test.

An application of this design would be an evaluation of the effects of mediation skills training on the acquisition of the desired skills. A pool of 48 trainees is assigned randomly to the four groups, 12 per group. Trainees in Groups A and C receive the training procedures; those in Groups B and D do not. A day before the training, Groups A and B are administered a pre-test that covers the kinds of mediation skills in the training package; Groups C and D do not get a pre-test. A day following the training, Groups A and C are given a post-test. Testing occurs at the same time, and with the same interval, for Groups B and D. Evidence for effective training would take the following form: Group A > (larger change in the direction of improved skills) Group B, Group C > Group D, and Group A (post-test) = Group C (post-test). However, these results do not provide evidence for the long-term effects of training. Such an evaluation would entail repeated post-tests given to each group through the course of several weeks or months. This design combines the features of the Solomon four-group design with before and after repeated measures. The form taken by this design is as follows.

Group A: R — O — O — O — X — O — O — O
Group B: R — O — O — O ——— O — O — O
Group C: R X — O — O — O
Group D: R O — O — O

The Solomon four-group design does not include a comparison of alternative training packages or other interventions. In order for such comparisons to be made, the design would have to be repeated for each of the several alternative procedures. Suppose that the training procedures (X1) used in the first design emphasized a settlement-oriented style (SOS) to mediation: According to Kressel and his colleagues (1994), this approach emphasizes the value of efficient, compromise solutions to conflicts. An alternative set of procedures, referred to as a problem-solving style (PSS), emphasizes the value of searching for information that can be used to reach an integrative (increased joint benefits) outcome. A question of interest is whether SOS (X1) is easier to learn than PSS (X2). The comparison would entail two Solomon four-group designs, one with SOS (X1) and the other with PSS (X2). An advantage of this comparison is that it controls for the threat to validity posed by testing effects. The key comparison would be the relative size of the differences between Groups A and B change scores (for X1 and X2) and between Groups C and D post-test scores. Larger differences for X1 than X2 would support the hypothesis that SOS skills are learned better than PSS skills. It would not indicate that SOS is a more effective approach to mediating disputes.

A disadvantage of comparing two or more Solomon four-group designs is the number of participants (or trainees) needed to fulfill the design's analytical requirements. The PSS (X2) design requires another 48 participants, bringing the total number of subjects for the experiment to 96. There is an alternative way of comparing the effects of different treatments. It is a short-cut with the attendant shortcomings. It consists of three groups, instead of the eight called for by the Solomon four-group design. A third group is a no-treatment (neither SOS nor PSS) control. Participating trainees are assigned randomly (R) to one of the following groups:

Group A: R X1 (SOS) ——————— O
Group B: R X2 (PSS) ——————— O
Group C: R X3 (control) ——————— O

If 12 trainees are assigned to each group (as in the designs above), only 36 participants are needed to implement the experiment. Support for the hypothesis that SOS is easier to learn than PSS would be indicated by higher post-test scores for SOS than for PSS and the control group (SOS > PSS = control); somewhat weaker support would also be obtained by the finding that SOS > PSS > control.

Although this design is easier to implement, it is a weaker evaluation of the hypothesis. When the pre-test is eliminated, no evaluation of changes resulting from the treatments exists. When a pre-test for each condition is installed, control for its effects independent of the treatments becomes necessary; hence, the preference for the Solomon four-group design. If, on the other hand, it is assumed that random assignment eliminates all pre-test differences (referred to also in the training literature as "baseline data"), then it follows that all post-test scores are deviations from the same pre-test result or baseline.

These are some of the considerations entertained by researchers torn between arguments that emphasize the value of reducing threats to internal validity at the cost of design efficiency (the Solomon four-group design) and those that promote design efficiency at the cost of dealing with alternative explanations for findings (the three-group comparison). It also gives the new researcher a glimpse into the way that decisions about research designs are made.

The mediation training example is an evaluation experiment that is concerned with impacts of mediation procedures on the development of skills. It focuses on outcomes rather than on the processes of skill acquisition. By paying attention to process, the researcher may gain insights into the way that the mediation procedures work. This sort of design would take the following form.

Group A: R X1 (SOS) ——— O1 (process coding) —— O2 (outcomes)

Group B: R X2 (PSS) ——— O1 (process coding) —— O2 (outcomes)

Group C: R X3 (control) —— O1 (process coding) —— O2 (outcomes)

Two types of assessments are made in this design, a coding of the process (O1) and outcomes (O2). The process may be captured by, first, observing the way trainees perform in role-play exercises during training and, second, developing codes that capture their progress. The outcomes are summary indicators of the mediation skills acquired. The analyses may include correlations between the process and outcome indicators, perhaps showing different paths (process) to acquisition (outcomes) for the different treatments. Some paths may hinder skill acquisition, whereas others may facilitate it.

The same type of experimental design was used in the simulation experiment by Druckman et al. (1988) discussed earlier. Recall that this experiment was a comparison of three conditions hypothesized to influence negotiation: interests derived directly from values (referred to as an "embedded" condition), values separated from interests (referred to as a "de-linked" condition), and a prenegotiation workshop intended to understand the contrasting values from which interests were derived (referred to as a "facilitation" condition). The post-test-only random-assignment design took the following form.

Group A: R X1 (embedded) ——— O1 (process) ——— O2 (outcomes)

Group B: R X2 (de-linked) ——— O1 (process) ——— O2 (outcomes)

Group C: R X3 (facilitation) ———O1 (process) ——— O2 (outcomes)

Equivalence—meaning that the average pre-test scores would be the same—among the conditions was assumed through random assignment. The embedded condition was hypothesized to generate the most competitive negotiations and, thus, regarded as the standard against which the other conditions were compared; the de-linked and facilitation conditions were hypothesized as variations in the direction of less competition. For these reasons, a pre-test was not necessary and would, in any event, be difficult to administer because the negotiation had not begun. No-treatment controls were also difficult to design because Groups B and C were variants on a theme much like a control condition would be. Process variables included codes for the types of statements made during the negotiation and self-reported perceptions. These variables intervened between the conditions and the outcomes (agreements reached, types of agreements, time to settlement).

The results showed that more agreements were reached in both Groups B and C than in Group A. However, these conditions (Groups B and C) differed in terms of process: more statements indicating agreement, more appeals to joint interests, and fewer statements indicating dominance by one negotiator over the other in the facilitation condition. These findings were interpreted as showing different paths to the same outcome, suggesting two models of negotiation, expedient bargaining for short-term agreements (in the de-linked condition) and cooperative bargaining for long-term relations (in the facilitation condition). The process paths are the "mechanisms" that connect the negotiating situation with outcomes. These results were generated from both the arrangement of the groups and the multiple assessments—processes, perceptions, outcomes—made possible by the experimental design. They contribute to the evaluation of interventions (see Chapter 11) and to theories of negotiation.

A way of bridging the two validities is by conducting field experiments. Two well-known examples from CA&R are the studies by McGillicuddy, Welton, and Pruitt (1987) and Mooradian and Druckman (1999). The former study was conducted in a New York state community dispute resolution clinic. This is a rare example of a field study that implemented a classical design by permitting random assignment of 36 disputant pairs and mediators to one of three experimental conditions: mediation without arbitration (med), mediators become arbitrators if the dispute is not settled (med/arb same), and a new party is appointed to be an arbitrator if the dispute is not settled (med/arb diff). A problem, however, was subject mortality: 68% of the participants dropped out of the study before it was concluded. Although this is a serious threat to the internal validity of the experiment, the investigators showed that the dropouts were not related systematically to the experimental conditions; the conditions did not differ in number of dropouts. Strong results were obtained on process but not outcomes: Med/arb (same) condition disputants were less hostile, made more new proposals and more concessions, agreed more often with the other disputant, and were more satisfied with the outcome than disputants in the other two third-party configurations. This condition induced more cooperative motivation than the other conditions (see also Pruitt, McGillicuddy, Welton, & Fry, 1989).

The Mooradian and Druckman (1999) quasi-experimental study was an attempt to compare the effectiveness of mediation (de-escalation hypothesis) versus a hurting stalemate (escalation hypothesis) in reducing violence between nations engaged in combat over Nagorno Karabakh. A large events data set was assembled (3,856 events during the period 1990–1995) and coded on a six-step scale ranging from most peaceful to most violent. These data were analyzed as a time series (dependent variables) influenced by the interventions—mediation or hurting stalemate (independent variables). The interventions were construed as experimental treatments in a sequence of before-and-after quasi-experimental designs without control groups. Significance tests (see Chapter 4) were used to assess the changes from before to after an intervention (mediation, hurting stalemate) in a manner similar to the way data would be analyzed from a laboratory experiment where control groups could also be included in the design. Results showed that the casualties suffered from the hurting stalemate reduced violence more than the attempts at mediation, leading to a cease-fire negotiation. The quasi-experimental design was useful for organizing comparisons in time (before and after interventions) and space (mediation vs. hurting stalemate).

For both studies, experimental designs contributed to knowledge about mediation. The internal validity threat incurred by dropouts in the study by McGillicuddy and associates (1987) was reduced by showing that the distribution was random across the conditions. The lack of random assignment and control groups in the Mooradian and Druckman (1999) study was offset to an extent by the number of opportunities to disconfirm the hurting stalemate hypothesis (six mediation efforts) and by the clarity of the definition of the stalemate (as offensives resulting in significant casualties; neither the casualties nor the mediators' behavior during the offensives were included in the events data set). Both studies are examples of opportunities to conduct well-designed experiments in the field. They are discussed further in Chapter 9.

Another approach to field experimentation is randomized field trials (RFTs). This approach avoids selection biases that occur in observational studies, including ethnographic research (see Chapter 8). However, RFTs have rarely been used in CA&R research to date. Researchers can benefit from the experience gained from RFT applications in other areas of social science. Several of these applications are discussed in the collection of articles edited by Green and Gerber (2004) for a special issue of the *American Behavioral Scientist.* That collection includes examples of field experiments on health behaviors, social welfare programs, criminology, and school vouchers. A particularly interesting methodological innovation discussed by Baruch et al. (2004) is random allocation of such macro-level units as villages and communities, housing developments, hospital units, schools, and other large administrative units. A good deal of conflict research is done at a macro level of analysis, with entities or places rather than individuals or small groups; very little of the research employs RFT (or place-randomized trials) procedures. The progress made to date in other areas bodes well for research that addresses conflicts between communities, organizations, or nations. These methods would also refine data collection strategies geared toward exploring connections between micro and macro levels of analysis. But a more popular approach to bridging the validities in CA&R research is simulation. We turn now to a discussion of simulation research and design with examples drawn from CA&R.

BRIDGING INTERNAL AND EXTERNAL VALIDITY: SOME SIMULATION DESIGNS

Simulation is another approach to experimentation intended to address threats to both internal and external validity. It is an attempt to represent a real-world context in the laboratory. The laboratory aspects of simulation consist of the experimental design. The contextual aspects of simulation consist of the scenario within which the experiment is embedded. A variety of randomized control group designs can be embedded in simulations. By employing these designs, a researcher reduces threats to internal validity.

Simulation experiments are also useful vehicles for exploring new conditions not present in a particular real-world setting. This feature is illustrated by the experiments discussed just below.

A variety of contexts have been simulated, although many simulations are over-simplified versions of those contexts. Examples are collective bargaining interactions, decision making in city councils, space missions, organizational departments, and diplomatic relations among national representatives in an international system. (Many other examples can be found in Crookall and Arai [1995], and Druckman [1990] as well as in the journal *Simulation & Gaming.*) To the extent that these simulated settings resemble the setting being simulated, the researcher reduces threats to external validity. A simulation is realistic to the extent that fundamental properties of the system (or context) being simulated are incorporated in the laboratory scenario. But this is also an empirical issue that can be evaluated by comparing findings obtained in the different settings. In making these comparisons, it is particularly important to specify clearly what is being compared (see Drabek & Haas, 1967). Several projects illustrate how this is done.

The most ambitious program designed to address issues of external validity is the research on simulated international processes conducted by Guetzkow and his colleagues. These researchers were able to collect comparable data from simulated and field settings. Their comparisons, made across more than 20 studies, showed considerable correspondence: 16 of 20 comparisons yielded judgments of much or some correspondence of findings between the settings (see Guetzkow & Valadez, 1981). Another interesting comparison of simulation and laboratory findings was made by Hopmann and Walcott (1977). They explored the relationship between external stress and negotiating behavior in a simulation of the Partial Nuclear Test Ban Treaty of 1963 and in the actual negotiations. In this study, the effects of three experimental conditions—benign, neutral, and malign systemic tension levels—were compared on a number of intervening process variables and dependent outcome variables. They also coded the negotiation transcripts on similar process and outcome variables. Variations in external stress as events unfolded during the course of the talks were tracked; this is different from creating experimental variations of stress. Convergences were found between the experimental and field results, indicating that stresses were dysfunctional for negotiations: High stress produced greater hostility, harder bargaining strategies, and fewer agreements than low stress. The experimental findings showed that the malign condition (high stress) produced significantly fewer solutions than both the benign (low stress) and neutral conditions. This kind of controlled comparison is a strength of experiments. It indicates that low stress does not improve the outcomes, but high stress hinders attempts to reach agreements.

A third study by Beriker and Druckman (1996) provides another example of the value of laboratory-field comparisons. These researchers compared findings obtained from a simulation of a post–World War I peace conference with the actual conference held at Lausanne, Switzerland in 1922–1923. A key feature of this conference was the power asymmetry among the parties. Symmetric (equal power) and asymmetric (unequal power) structures existed on two important issues in the talks, passage to the Black Sea through the straits and the question of civil rights for minorities. A content analysis of the transcripts showed some differences in bargaining behavior between the two power structures.

These structures were then simulated and compared to a third condition, bilateral negotiation between relatively weak parties of equal power. The additional condition created for the simulation provided insights not likely to be obtained from an analysis of the actual talks. Negotiators in the bilateral condition were more satisfied with the outcome, achieved faster resolutions, disagreed less, and made fewer competitive statements during the discussions than those in the multilateral (coalition) conditions. Similar to the Hopmann and Walcott (1977) simulation, this study illuminates an advantage of experiments: They allow for exploration of new conditions not present in the setting being simulated; these explorations often produce interesting findings with practical implications. The comparison between the two power-configuration conditions that existed in both simulation and field showed some similarities and some differences in the findings. The similarities bolster support for the relevance (or external validity) of the simulation to the case being simulated.

External validity can also be understood in terms of how much complexity is captured in the simulated experimental environment. Most field settings are more complex than experimental settings, and many conceptualizations of conflict resolution processes include more factors or variables than can usually be investigated in an experiment. Thus, experiments are usually only limited evaluations of these frameworks.

An example is a study that attempted to evaluate the framework constructed by Sawyer and Guetzkow (1965) to encompass a variety of processes and influences of negotiation. This framework depicts pre-conditions, background factors, conditions, processes, and outcomes of negotiation (see Chapter 2). The challenge is to evaluate the impacts of these multiple interacting influences on negotiating behavior. This was done by constructing "packages" of variables in a simulation scenario of a multilateral environmental conference (Druckman, 1993).

Three packages were developed for each of four stages of the conference: *pre-negotiation planning, setting the stage, the give-and-take,* and *the endgame.* Each of the packages contained five or six variables hypothesized to influence the flexibility of the negotiators. The design enabled me to compare the effects of three different types of packages at each of the stages: one in which the variables in all the stages were arranged to encourage flexibility, another where they were arranged to encourage inflexibility, and a third where the variables in the earlier stages were arranged for flexibility and those in later stages toward inflexibility. The three packages were regarded as experimental conditions. This was a 3×4 design in which the three conditions were combined with the four stages (a repeated measure).

Effects of these conditions were evaluated for their impacts on the flexibility of the negotiators in the simulation. However, this evaluation did not reveal the relative importance of the variables contained within the packages. For this evaluation, a procedure known as a "halo error" technique (Guilford, 1954), was used to unpack the variables, providing weights that indicated the relative impact of the variables on flexibility within the stages. This experiment is a good example of how simulation can be used as a device to evaluate the sorts of combinations of variables depicted by frameworks. It shows how complexity can be incorporated in experiments. To the extent that this goal is achieved, external validity is enhanced.

Simulations are a way of addressing issues of external validity raised by laboratory experiments. In this section, I have shown how some researchers evaluate the correspondences between findings obtained in simulation experiments and in the settings being simulated. To the extent that correspondences are found, a case can be made that the simulation is an accurate model of the real-world setting. The strong correspondences found for both international decision making and negotiation processes bolster confidence in the external validity of the simulations. This does not mean that other simulations are also good replicas of the settings being simulated; the evaluations must be done separately for each simulation. However, some design guidelines intended to enhance external validity can be suggested.

One guideline is to incorporate complexity in the simulation, as illustrated by the environmental conference discussed above. Another is to recruit participants for the simulation from the real-world setting, for example, diplomats for a study of international negotiation or scientists for a study of environmental policymaking. A third is to perform multi-method research by analyzing data collected from both the simulation and field settings. The field experiences are likely to suggest structures and variables that would improve the validity of the simulation; the simulation may contribute methods of analysis useful for the field study as well as evaluate the impacts of new situations not readily available for analysis in the real world. For more reading on the issues and challenges of evaluating a large variety of types of simulation validities, see Feinstein and Cannon (2002). One interesting distinction made by these authors is between the structure of the game and the effects it has on those who play it. The former is referred to as representational validity, the latter as educational validity. The questions asked are as follows: How similar are the simulated and referent (real-world) environments? Do participants in both environments behave in similar ways? Both are important questions.

Simulations as Models

The scenarios discussed above are attempts to represent complex activities or social systems. To the extent that they succeed in doing this, the simulation is regarded as being externally valid. The research reviewed in the previous section shows that one way to evaluate success is to compare results obtained in simulated and real-world environments. Judgments of external validity turn on the similarity of results.

Another way of thinking about this issue is in terms of simulation design. With Robert Mahoney, I proposed a taxonomy for distinguishing among types of simulation designs (Mahoney & Druckman, 1975). It consists of two dimensions referred to as range and extension. *Range* refers to the number or variety of situations that are modeled. Simulations that are applicable only to a few well-defined situations are said to cover a narrow range. Those that represent many types of situations cover a wide range. *Extension* refers to the amount of detail incorporated in the simulated environment. When many details are written into the scenarios, the simulation is said to have deep extension; when few details are included, the simulation has thin extension. The simulations of international processes, discussed above, cover a narrow range and have deep extension. The popular prisoner's dilemma game (PDG), on the other hand, is a wide-range, thin-extension simulation. Both of these types of simulations can be thought about as models. Examples of these models are presented in this section, focusing first on game-theoretic models, second on computer simulation models, and third on more complex computer and human simulations on problems of conflict.

GAME-THEORETIC MODELS

Game-theoretic models are parsimonious in the sense that they derive their power from simplicity. Osborne (2004) notes that "the assumptions upon which (game-theory models) rest should capture the essence of the situation, not irrelevant details" (p. 1). The key assumption is based on the theory of rational choice: It posits that a decision maker chooses the best action according to his or her preferences among all the available actions. Preferences are often represented by payoff functions, as shown in the examples below. But there is another important assumption based on the idea of interdependent decision making: It posits that the best action for any given player depends on the other players' actions. This means that a player must have in mind (or form a belief of) the actions that the other players will take. These assumptions are the underlying bases for the most popular solution concept, which is referred to as a *Nash equilibrium.*

Nash (1950) showed that a unique solution to the bargaining problem can be identified for all two-person games. The characteristics of this solution can be summarized simply as the outcome that maximizes the product of the preferences (utilities) of the two parties. Expressed differently, this solution minimizes the losses incurred by players who choose other outcomes in a game. As such, it may be regarded as a steady state or a social norm in the sense that if everyone else accepts it, no person prefers to deviate from it. Although other solution concepts have been proposed, the simplicity and elegance of the Nash concept has given it the longest "shelf life." (For a review of various solution concepts and the studies designed to evaluate them, see Schellenberg & Druckman, 1986.)

An advantage of game-theoretic analysis is that conclusions or solution concepts are derived from general models that cover a wide range of choice dilemma situations, in which a person must decide between making an offer or remaining silent, voting for one candidate or another, or intervening versus standing aside. It does not depend on knowledge of particular features of the specific environment in which actions are taken. (Recall the distinction made between etic and emic approaches to research in Chapter 1.) For example, social-psychological analyses of bystander apathy attempted to understand the circumstances in which a bystander would help a person in trouble. The analyses revealed that group size mattered, and the investigators posited alternative hypotheses to explain why there is a decline in offering assistance as the number of witnesses increase: Alternative explanations include diffusion of responsibility, audience inhibition, and social influence. A more parsimonious explanation is based on the idea of an equilibrium. Following Osborne (2004),

> Whether any person intervenes depends on the probability she assigns to some other person's intervening. In an equilibrium each person may be indifferent between intervening and not intervening, and . . . this condition leads inexorably to the conclusion that an increase in group size reduces the probability that at least one person intervenes. (pp. 133–134)

This conclusion is based on fewer assumptions and follows logically from them. These abstract concepts may come to life in the illustrations of various types of games to follow.

The prisoner's dilemma is a two-person (or player) game that models the tension that often exists between trust and risk. In order to achieve the best outcome or highest payoff, players must trust each other to cooperate rather than to compete in an exploitative fashion. The best individual outcomes result from playing the game competitively; the best joint outcomes occur when the players choose cooperatively. This dilemma is reflected in the game matrix presented to the players. It takes the form illustrated in Figure 3.1.

The numbers represent payoffs to each player, Suspect 1 and Suspect 2. The highest payoff is 4; the lowest is 1. In terms of the game story, an outcome of 4 means no time spent in prison; an outcome of 1 means a stiff sentence with lighter sentences imposed for the intermediate outcomes of 2 and 3. It can be seen that each player (suspect) receives the highest payoff (4) when he or she confesses and the other remains silent. The highest joint payoff (known as the optimal outcome) occurs when both remain silent (total of 6), the lowest when both confess (total of 4). The idea of trust is represented in the game by a prediction that the other will remain silent. The idea of risk is reflected in acting on this prediction by remaining silent. If the other thwarts my prediction by confessing, I receive the lowest payoff of 1 (maximum prison sentence) while he or she gets the highest payoff of 4 (or freedom). This is the dilemma confronting

		Suspect 2	
		Remain silent	Confess
Suspect 1	Remain silent	3, 3	1, 4
	Confess	4, 1	2, 2

Figure 3.1 The Prisoner's Dilemma Game

both players. Most outcomes of PDG experiments are confess-confess (a total payoff of 4). Players tend to be risk averse or competitive rather than trusting. This is also known as the equilibrium outcome, defined as the outcome that minimizes the players' losses rather than maximizes their gains.

The reason why this is regarded as a wide-range, thin-extension model is that it captures a dilemma present in a variety of conflict situations. It highlights the mixed motives of cooperating and competing found in many negotiating situations. This dilemma is reflected in a simple matrix form, without the need for elaborate substantive information about issues, history, and situations. For this reason, however, the PDG does not capture the complexity of detail found in any particular conflict situation. Thus, it differs from the examples of role-play simulations, such as collective bargaining, discussed in the section above. It is one of a family of matrix games, each of which illuminates a particular dilemma faced by parties in conflict. These games include chicken, deadlock, bully, maximizing differences, coordination, and battle of the sexes (for a lucid presentation of these games and their applications to actual international conflicts, see Snyder & Diesing, 1977).

The game of chicken takes the form illustrated in Figure 3.2. In this game, the best outcome for each of the players is obtained when he or she does not swerve and the other swerves (4). The best joint outcome occurs (each gets 3) when both swerve, and the worst outcome occurs, of course, when neither swerves. Similar to the PDG, the dilemma is between trust (predicting that the other will swerve) and risk (acting on the prediction by not swerving). A difference between the games, however, is that the consequences are more severe for making the wrong prediction in chicken: Collision occurs. Another difference is that the equilibrium solution is the swerve-swerve outcome in the upper left box rather than the confess-confess outcome in PDG's lower right box. This format has been used to capture the interactions between President Kennedy and Premier Khruschev in the famous historical situation of the Cuban Missile Crisis. A question raised by historians is whether the outcome

		Player 2	
		Swerve	Do Not Swerve
Player 1	Swerve	3, 3	2, 4
	Do Not Swerve	4, 2	1, 1

Figure 3.2 The Chicken Game

of that crisis was the equilibrium solution, swerve-swerve (both removed missiles; the United States from Turkey, the Soviet Union from Cuba), or the solution favoring the United States, swerve-do not swerve (Soviets remove missiles as United States stands firm). Another historical situation that has been modeled as a game of chicken is the Berlin Blockade of 1948–1949. The outcome of this crisis was swerve (the Soviet Union conceded)—do not swerve (the United States held firm) (see Snyder & Diesing, 1977, p. 114).

The maximizing-differences game takes the form shown in Figure 3.3. This is an interesting game from the standpoint of revealing the social motivation of players. The "rational" choice for both players is A: Both maximize their payoffs with this choice. Yet, despite the obvious, many players choose B, not because they do not understand how payoffs result from choices, but because they prefer a solution that maximizes the difference between them for relative gain rather than obtain the highest individual and joint payoff for absolute gain. McClintock and Nuttin (1969) showed that the frequency of relative gains choices differs for Dutch and American children. Hopmann (1995) discusses the prevalence of a relative-gains approach taken by nations in foreign policy decision making and negotiations, particularly during the cold war.

		Player 2	
		Choice A	Choice B
Player 1	Choice A	6, 6	0, 5
	Choice B	5, 0	0, 0

Figure 3.3 The Maximizing-Differences Game

		Player 2	
		Choice A	Choice B
Player 1	Choice A	2, 3	1, 4
	Choice B	4, 2	3, 1

Figure 3.4 The Bully Game

The payoff configuration for the bully game is as shown in Figure 3.4. This game illuminates a dominant choice for Player 1. This player can never lose if he or she chooses B: The difference in payoffs is between 4 and 3, depending on whether Player 2 chooses A (1 gets 4 units) or B (1 gets 3 units). This game depicts interactions between parties with different amounts of power. It is particularly relevant to negotiations between hegemonic powers and lesser powers who depend on them. An example from ancient diplomacy involved negotiations between the pharaohs of Egypt and emissaries from lesser kingdoms, as illustrated by Druckman and Guner (2000). (See also Guner & Druckman, 2000, for further elaboration of the model.) It is also illustrated by Leng (1998) in terms of influence strategies used by 20th-century powers in recurring crises between pairs of nations.

This family of games models relatively simple conflicts. They can be compared to one another in terms of the symbolic notation given in Figure 3.5. In this notation, R is reward, T is temptation, S is a sucker's choice, and P is punishment. The sequence of outcomes for the PDG, ranging from best to worst for each player, is $T > R > P > S$. This means that the highest payoffs go to Player 2 in the upper right-hand box; the highest payoffs go to Player 1 in the lower left-hand box. These payoffs result from the combined choices of both players. The sequence of outcomes for chicken is $T > R > S > P$. In this game, the punishment is quite severe and worse than making the sucker's choice to swerve when the other does not. Notice the difference between these games: S and P are reversed, with S being better (worse) than P in chicken (PDG). Each of the other games is defined by a sequence of these outcomes. For example, the sequence for the game of deadlock is $T > P > R > S$: PP (in the lower, right-hand box) is the equilibrium outcome.

The configuration of outcomes in matrix games has a strong influence on the way that the games are played. In most games, however, the best joint outcome (which is often the R, R choice combination) is not the same as the

	Player 2	
Player 1	R, R	S, T
	T, S	P, P

Figure 3.5 Symbolic Representation of Game Payoffs

best individual outcome (which is often the S, T or T, S choice combination). Thus, advice would be tailored to game-playing goals that can be competitive/individualistic or cooperative/collective. These goals are easily manipulated through instructions about how the game is to be played. But there are a number of other variables shown in experiments to influence outcomes. The long list includes the size of payoffs, whether choices are made sequentially or simultaneously, whether feedback is provided, the number of games, the relationships between players, framing of the payoffs as gains or losses, whether the player roles are representative of others, and so on. This has been a popular line of research due, at least in part, to the ease with which experiments are conducted and embedded in very simple scenarios. It is also a flexible form of modeling that reduces a large variety of real-world conflicts to their essentials. (In addition to the 20th-century applications given in Snyder & Diesing, 1977, see Guner & Druckman, 2000, for analyses of four games representing different assumptions about information exchanged in diplomatic correspondences between representatives of the kingdoms living during the Bronze Age [circa 1400 BCE].)

Game-theoretic analyses have also been used to demonstrate some counterintuitive aspects of the mediator role. In his analysis of mediator bias, Kydd (2003) showed that communications may be more credible when the mediator is biased toward one or another disputing party. In bargaining situations, where the interests of the parties are at least partially opposed, there is an incentive to bluff: A party is encouraged to tell the mediator that it has high resolve and will fight unless it receives a concession, regardless of whether or not this is the case. The question then is under what conditions can a mediator credibly communicate information to the negotiators about the other party's resolve? Much of the literature on mediation promotes the idea that an impartial mediator can be trusted by all parties because he or she neither favors nor disfavors any party.

Kydd's (2003) game-theoretic model demonstrates the opposite of this conventional wisdom—that a partial mediator is likely to be a more effective communicator. The model implies that a mediator who favors the side he or she is communicating with is more credible, and thus persuasive, about the other side's resolve. This then encourages the favored party to make a concession. The model also shows that a certain degree of bias is not only acceptable but actually necessary in some roles the mediators play and that the function of providing information can be implemented better, in some circumstances, by powerful, biased mediators. (For a readable introduction to the many facets of game theory, including games with perfect and imperfect information as well as variants and extensions, see Osborne, 2004.)

COMPUTER SIMULATION

Another form of modeling is computer simulation. Instead of having people play games in assigned roles, the computer performs computations based on assumptions specified in a model. The assumptions may, for example, consist of preferences for alternative possible agreements that would settle conflicting claims between two or more parties. The preferences are often expressed as probabilities assigned to each proposal or as a probability distribution across the complete set of possible proposals. In an interactive context, the probabilities or preferences determine the proposals (referred to also as demands) made by each of the disputing parties. Agreements or impasses emerge from sequences of proposals and counterproposals. The key to agreement is the extent to which the parties change their proposals or make concessions. According to one model, concessions are a function of three parameters: desire to reciprocate, desire to initiate reciprocation, and friendly feelings toward the opponent.

Bartos (1995) expresses these parameters in the form of an equation:

$$dD = -k * dO - a * D + g \quad 0 < k, a < 1 \tag{1}$$

where D is the current demand, dD is the current change in demand, and dO is the current change in the opponent's offers. Parameter k controls the tendency to reciprocate; a, the tendency to make unilateral concessions; and g, the level of feelings. This model specifies the conditions under which concessions will be made: The larger the reciprocation parameter (k), the larger a party's concession will be in response to an opponent's concession; unilateral concession making (a) is most pronounced during the early stages of negotiation when demands are high, and the larger the friendliness or feelings parameter (g), the larger the concession. The model is evaluated by varying the values of these parameters and observing when an (and what) agreement is reached. Bartos (1995) used this equation to determine paths of demands: The computer generated a series of demands that, according to the probabilities specified by the program, each negotiating party would make. He then evaluated the relative merits of two approaches to negotiation, a *distributive* (concession exchange) and *integrative* (information search) approach. He concluded that the distributive process is faster, but the integrative process may be more productive because it can increase the chances of an agreement.

A second example of computer modeling involves stochastic processes. These processes assume sequential dependence of events. A number of social conflict processes have been modeled as a continuous-time Markov chain. This model assumes that the outcome of one event places an individual in a particular state, and the probabilities of new events depend on that state. It construes conflict as a dynamic process in which parties transition between states of antagonism and cooperation. A key analytical question is, What influences the rates of transition from one state to another? Earlier work by Coleman (1973) on causal modeling made progress on this issue. He showed that transition rates are a function of two parameters expressed by the following equation:

$$q(t) = q_0 e^{-at} \quad (2)$$

where q_0 is the transition rate at the time of the initiating event, a is the rate of decline, and $q(t)$ is the transition rate at time t after the initiating event. This equation captures exponential decay in rates of change from one state to another. Decline in rates is a function of time since the last move: The longer a process remains in a state of cooperation between parties, the lower the probability that it will move to a state of antagonism between them. Empirical examples given by Coleman (1973) illustrate that the longer parties remain in a state, the less they attend to cues that may signal a change in process or relationship. However, this process is reversed when there is a terminating event such as an election or a negotiating deadline. In this case, transition rates increase over time in a systematic fashion. The exponential growth function that captures this process takes the following form:

$$q(t) = q(0)e^{at} \quad (3)$$

where $q(0)$ is the value of q when $t = 0$, and the other terms are same as that in Equation 2.

These two situations of initiation and termination can be studied with computer simulation or experimental methods. One example would be to manipulate the time spent in each of several stages of a conflict process, with and without an imposed deadline. Calculations of rates of change (q_0 and $q(t)$) would be made for the same time period when a deadline has been imposed and when it has not. Simulation data showing that rates slow down with time from initiating events and speed up with time to a terminating event would have implications for third-party interventions intended to move disputing parties out of antagonistic states. Creating a terminating event can move the parties quickly out of a state toward agreements. When terminating events are not available or cannot be manufactured, early interventions that are less direct or more facilitative may work. Later interventions may work better if they are more direct, including the use of combinations of mediation and arbitration (see McGillicuddy et al., 1987).

An application of stochastic modeling to international interactions was developed by Duncan and Job (1980). By using coded data on tensions, they calculated transition probabilities—modeled as a continuous-time Markov process—for estimating shifts from one state (minor tension) to another state (major tension) of the interaction system. The probabilities provided additional information about trends in the interactions between Israel and Syria and between Rhodesia and Zimbabwe. The information is useful for forecasting when changes are likely to occur, contributing to the technology of early warning of escalations or de-escalations in the ongoing relationships.

The models discussed to this point in the section are relatively simple. They deal with only a few variables or parameters: Demands, offers, reciprocation, and feelings in Bartos's bargaining model (1995); time from initiating events and to terminating events in Coleman's and (1973) model of transition rates. The game-theory models focus on two choices made by each of two players confronted with a dilemma of balancing trust against risk. The relationships specified by these models can be evaluated with two-person game experiments conducted over a brief period of time. They can also be evaluated in terms of assumptions about expected payoffs to the players. An example of this kind of strategic analysis is Sandler and Arce M.'s (2003) game-theoretic treatment of terrorism. They developed alternative scenarios, based on different types of games (e.g., PDG, coordination), for assessing the benefits of proactive or reactive government policies. They also explored some implications of granting concessions to terrorist organizations with differing proportions (p and $1 - p$) of moderate and hard-line members. These implications include answers to the following question: What kind of offer should a government make to a terrorist organization? By addressing these issues, the models provide a means for evaluating the effectiveness of antiterrorist policies.

COMPLEX MODELS

More complex models are needed when the number of conflicting parties in a system increases (e.g., Mosher, 2003) or when general variables (such as terrorist incidents) are decomposed into their various components (such as intensity and location of incidents, type as own or foreign terrorism, effects on market share and spillover; e.g., Drakos & Kutan, 2003). These kinds of models are often tested with archival data over time. They can also be evaluated with simulated data, as illustrated by Clarke's (2003) attempt to discriminate among different models of international relations. (He performed Monte Carlo experiments with randomly generated simulation data.) Further opportunities for simulation modeling are provided by Internet technologies. Dasgupta (2003) provides an example of the use of Internet-mediated simulations for providing uncontrolled environments for the study of group decision-making processes. For a survey of various applications of Internet technologies, see the other articles in the special issue of *Simulation & Gaming* edited by Dasgupta (2003b).

As well, data collected from human role-playing simulations can be used to evaluate complex models. Typically, however, the models evaluated by role-play simulations take the form of schematic diagrams or flowcharts rather than sets of equations. Examples include Wolfe's (1995) evaluation of the management development process; the Klabbers, Swart, Van Ulden, and Vellinga (1995) use of gaming to evaluate problem-solving models for addressing climate change and policy; and the Garris, Ahlers, and Driskell (2002) use of instructional games to demonstrate the value of an input-process-outcome model of learning. In addition to model precision and calibration or measurability, the computer and human forms of simulation differ in terms of the number of variables in the model and the consistency of the behavior observed. Computer simulations—especially those based on mathematical models—deal with fewer variables and focus attention primarily on the rational or ordered aspects of decisions or behavior (see Armstrong, 1995, for a taxonomy of simulation approaches).

Of particular interest in this discussion of simulation approaches to modeling is the trade-off between broad and narrow applicability of computer or role-play results. The broad applicability of game models is gained at a price of relevance to particular conflict situations. Relevance is less at issue when simulations are used primarily as vehicles for theory development, as in the examples of the Bartos (1995) and Coleman (1973) models. It is more important when simulations are used to accomplish training or policy goals, as in the example of climate-change policy or instructional simulations. In Chapter 1, I discussed the difference between abstract or general and concrete or specific concepts. The game models—and other wide-range thin-extension approaches—are closer to the abstract end of this continuum. Many policymaking simulations, as examples of narrow-range, deep-extension models, are closer to the more concrete end. Abstractions are useful for theory construction. They are less useful for giving advice about particular issues or situations. They may also be less useful for participant learning about how to perform well-defined tasks in other settings. (For more on issues of transfer of learning, see Reder & Klatsky, 1994; on types of learning, see Kolb, 1984.)

Summary

The discussion in this chapter reviews the variety of designs that have been used in experiments. Examples of experiments that address CA&R issues are used to illustrate the designs, which vary in complexity from simple pre-post test comparisons to elaborate controls with repeated before-and-after treatment assessments. The key idea for experiments is random assignment of participants to conditions. This idea is the primary distinction between quasi and classical experiments. It bolsters confidence in the internal validity of an experiment, especially when efforts are made to control for various possible threats to validity. It does not ensure external validity, however. Two ways of bridging internal with external validity is through field experimentation and simulation design. The advantages of randomized field trials have not been exploited by CA&R researchers to date. Simulation, on the other hand, has been a popular approach to doing research and modeling.

Simulation designs are attempts to reproduce real-world environments, including historical negotiations, political decision making, and organizational processes. To the extent that they are effective in doing this, external validity is improved. The examples given show how experiments can be embedded in simulations of conflict. Simulations are also models of social processes. The models discussed include both relatively simple games such as the popular prisoner's dilemma and the more complex designs that capture many details of the settings being simulated. The former are thought to apply to a wide range of situations and, thus, may be useful devices for theory testing and development. The latter models, which include both computer and human simulations, focus on particular situations and may thus be useful for practice and training.

In this discussion of experimental designs, methods of analysis have been mentioned but not developed. Let us turn to the next chapter to consider how data collected from experiments and simulations can be analyzed.

Discussion Questions

It is hoped that you now have an appreciation for both the strengths and limitations of experimentation. The strengths may lead you to consider performing experiments on CA&R topics. The limitations should lead to a consideration of using a multi-method research strategy that includes doing experiments. Several discussion questions are suggested by this presentation of the experimental and modeling approaches to research. You may want to use them as a basis for class discussion and review before moving on to the next chapter on methods of analysis.

1. What is the primary goal of experiments? What are the key features of this approach to doing research?
2. Internal and external validity are often regarded as being complementary considerations in experimental design: Give examples of how you might attempt to reduce some threats to both types of validity. What are some trade-offs between the two types of validity?
3. What are the key distinguishing features of pre-experimental, quasi-experimental, and classical experimental designs? Under what conditions would one or the other of these types of designs be used?
4. Design a CA&R experiment that combines the Solomon four-group design with time-series data. What are some advantages of using this type of design compared, for example, to a one-shot, pre-test–post-test design?
5. Many experiments are designed to explore the connection between the experimental variable or treatment and outcomes. What other kinds of data might be useful for explaining this connection? Give examples from studies of conflict resolution.
6. How can experimental simulations be used to bridge the gap between internal and external validity? What are some strengths and limitations of simulation techniques?
7. What is the relevance of the distinction between models that have wide range and thin extension and those that have narrow range and deep extension? Give examples of models in each of these categories.
8. Describe some contributions made by game-theoretic analyses to the field of CA&R. Illustrate these contributions with games such as prisoner's dilemma, chicken, or bully.
9. What is the value of computer simulation approaches to understanding (theory) and resolving (practice) conflicts? What are some gains (and some losses) from using computer platforms rather than human role players in modeling?
10. Compare relatively simple (a few parameters) with complex (many parameters) simulation models. What are the relative advantages and disadvantages?