FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 96–106

ISSN 2081–6642

Jane Sell
Texas A & M University, USA

Murray Webster, Jr.
University of North Carolina – Charlotte, USA

The Importance of Cross-Cultural Experiments for the Social Sciences

Abstract

We detail how experiments differ from other types of research. In particular, laboratory experiments involve random assignment and control. They create artificial environments that are designed to test particular kinds of theories – those composed of concepts not bound to space and time. The aim of cross-cultural experiments is to ensure that theories are not tested in one cultural setting. Such experiments can help recognize how different cultures uniquely identify different initial conditions for theory development.

Key words: laboratory experiments, cross-cultural research, random assignment, initial conditions

Experiments and their characteristics

Cross-cultural research helps us find out whether and how culture affects basic principles of human behaviour. In this paper we examine some of the fundamental purposes and properties of experiments in order to determine when experimental cross-cultural research is most useful, when it is less useful, and how it can inform both theory and description.

There are many different methods of social science investigation: documentary-historical, survey, participant observation, and experiments. All of these methods can be used to test theories but they vary in the way in which the theory testing is accomplished. *Experiments* differ from other methods in several ways. First, the investigator by means of experimental 'manipulation' can endow the units of analysis (individuals or groups) with varying levels of the independent variable(s) to study the effects that inter-level differences are predicted to have on the dependent variable(s). Other methods do not create changes in independent variable(s), relying instead on observation or measurement in some settings. In the case of laboratory experiments, manipulation occurs within the artificial setting of a laboratory. In the case of field experiments, it can occur in different natural contexts such as schools, hospitals, and organizations.

Second, experiments utilize *random assignment* of treatments or conditions to the units of analysis, usually, people or groups. Randomization is an extremely important aspect of any kind of method as it serves to eliminate bias in the data. It often

takes the form of *random sampling* of respondents for surveys from some population or random sampling of articles for content analysis. In random assignment, there is nothing random about the participants; in fact, experimental subjects are often recruited for a given study because of having a particular characteristic. There is nothing random about cancer patients being recruited for a new cancer drug study, or university students being recruited for a study about learning techniques. The random assignment of treatments to individuals or groups is the random component of experimental method and its distinguishing mark. For example, the assignment of a placebo vs. a new experimental drug to cancer patients or the assignment of a visual vs. audio learning technique to students is done so that each patient or student has an equal (or known) probability of receiving each of the two treatments.

Because laboratory experiments are conducted in controlled conditions in which random assignment is feasible, they are *artificial*. Such artificiality also enables control or holding certain conditions constant. Control and randomization are powerful ways to eliminate alternative interpretations for the results of the experiment. Basically, this means that if a researcher has carefully designed the experiment, he or she can be certain that the outcome (a change in values of the dependent variable) occurs due to the independent variable.

While randomization eliminates an infinite number of alternatives for the results (such as irrelevant characteristics of individuals and what people are wearing), control eliminates alternatives associated with factors researchers know to make a difference. As an example, in experimental studies of group interaction, we know that the history of a group and of its members makes a difference. Consequently, we control on that history – often by creating groups that did not have prior interactions. As another example, we know that the information that the group members have about one another can dramatically affect the interaction process. When a person knows that the people with whom he or she is interacting share certain status characteristics, the interaction can be very different from that in the situation where the group members share none of the same status characteristics.

To eliminate alternative interpretations of the results, researchers must have a well-formed theory that pertains to the phenomenon/a under study. If the theory under investigation concerns the development of a lasting group dynamic that was set in motion by the very first interaction, the research design should enable defining the initial state of the process, as well as a series of interactions to allow the measurement of stability. At a very elementary level, this illustrates how a theoretical question must be carefully articulated before any design is considered as a way to answer the question.

Experiments can test particular types of relationships, which are expressed by *theoretical principles* of a particular kind, namely, those independent of space and time. This means that the concepts or terms used in these principles and thus the principles themselves are 'exact class.' The term *exact class* is taken from the logical distinction between concepts that are defined in such a way that their meaning never changes, and those that are 'ordinary' or changeable and malleable. This distinction has been articulated by Stephen Korner (1966) as a means to distinguish different approaches toward knowledge. Exact class concepts obey the law

of excluded middle because they are defined precisely so that it is possible to determine whether an event or object is or is not a member of a class of events. It either is or is not. There is no 'middle,' or 'almost.' Additionally, the meaning of such concepts never changes. Mathematical concepts such as 'triangle' provide a good example. A triangle is defined precisely and one can always determine if something is or is not a triangle. A square is not a triangle, in part, because it has more than three sides. The concept of triangle is the same in 2016 as it was in 1942 or 1900. No particular time or location is necessary for determining if an object is or is not a triangle. Hypotheses or propositions that incorporate triangles can be replicated precisely because they are not defined by a specific time.

On the other hand, concepts such as World War II, gender roles in 2016, and the Paris Climate Accord are ordinary terms. Their definition is absolutely conditioned by certain points in time and they are only understood in historical context. While we can, most certainly, study these concepts, we cannot replicate them because they are uniquely situated in history. Propositions or questions that utilize ordinary concepts are meant to capture history or historical change, and so are descriptive. Questions such as, 'what are Americans' attitudes toward Donald Trump in July of 2016,' are descriptive because they incorporate ordinary concepts.

While descriptive questions are critical for developing understanding, they are not appropriate for experimental tests. Experiments are artificial and as such, are poor instruments for studying particular events or even change. We would not bring large groups of Americans into an artificial setting to ask about their attitudes toward different political figures – there would be no need. This is a descriptive question. Random assignment or experimental control would be inappropriate. However, if we were interested in a snapshot of how Americans felt about political figures, such as Donald Trump, at a particular point in time, we would be concerned about *random sampling* to ensure that we could generalize to the existing population at this particular point in time. Experiments, on the other hand, can test principles utilizing concepts that do not lose their meaning in a certain context or time: concepts such as status, evaluations, public goods, task success, and behavioural constraints. These concepts, if defined exactly, can be used to examine events in the past or even in the future.

Scope conditions and initial conditions

The principles or propositions composed of theoretical or exact class concepts must be accompanied by *scope conditions*, or the parameters that delineate when propositions or hypotheses are expected to obtain. All theories have limited scope. No theory applies all of the time, and the scope conditions delineate, abstractly, when the theory being investigated is thought to apply. Newton's famous laws of acceleration of falling bodies apply only in the absence of factors such as air resistance or magnetic deflection; in other words, they take as scope conditions the absence of those factors. Expectation states theories, a prominent social psychological set of theories, describe how people in task groups organize interactions. Many of the formulations

delineate scope conditions requiring that task-oriented groups in which people meet have no history of interaction. Scope conditions are part of the theoretical specification of the theory or sets of propositions (See Cohen 2003; Walker, Cohen 1985; Foschi 1997; Sell, Martin 1983; Webster, Sell 2014). They too, are composed of concepts that are abstract because they are part of the theoretical formulation or foundation. They 'travel' with the derivations. So, if we were discussing expectation states propositions that involved scope conditions of task-oriented groups with no prior history of interaction, these conditions would follow through all the propositions and derivations of the theory.

It can be noticed that the theoretical principles mentioned are vastly different from descriptive principles that specify times and places. Because of this, these theoretical principles are artificial in the sense that they are not of our experience. Because these concepts and principles are artificial, the artificiality of the laboratory setting makes it ideal for testing. In this way, artificiality of experiments is not a disadvantage as it is sometimes mistakenly thought to be, but rather a distinct advantage, at least for dealing with theoretical principles. Other methods are not as artificial because they deal with complex everyday settings, such as actual organizations, institutions and groups in which variables cannot be easily disentangled and, of course, cannot be manipulated. Theory testing is certainly possible with the use of other methods, but causal mechanisms, in particular, are more difficult to disentangle than with the use of experimental methods.

At the same time, there must be a translation of the theoretical concepts to concepts that can be used in testable hypotheses. This process, termed *instantiation*, means that the theoretical concepts must be defined in ways that are measureable in a specific setting, in a specific place. This requires the experimenters (or other researchers as well) to develop what are usually termed 'operational' measures of abstract concepts.

Instantiations of scope conditions are termed *initial conditions*. While theoretical principles and scope conditions do not describe specific settings like a group of college students in Cracow, they certainly can apply to such settings. The key to such application is the use of initial conditions (Cohen, 1989, 2003; Foschi 1980, 1997; Webster, Sell 2014). They are specific to the setting and supply a 'starting place' for the testing of the principle. Because this is the case, the researcher must understand the cultural setting. For example, if we want to investigate how information about diffuse status affects performance within a group, we need to know what a diffuse status characteristic is in the particular culture (and the particular time) we are going to study. Or, alternatively, we could create a diffuse status characteristic, a topic which has already been investigated in a number of recent theoretical explorations (Webster, Hysom 1998; Ridgeway, Correll 2006; Ridgeway et al. 1998; Ridgeway et al. 2009; Ridgeway, Erickson 2000). A *diffuse status characteristic*, as defined within status characteristics theory, is a characteristic of an individual with at least two states differentially evaluated. With each state there are associated specific performance expectations, as well as general performance expectations. That is, those who possess the higher state of this diffuse status characteristic not only are societally defined as better at task performance than those who possess

the lower state of the characteristic, but those individuals are perceived as 'generally better' than those who have the lower state of the diffuse status characteristic. This term is exact class because it does not refer to specific time and place and is precisely defined so that it is possible to determine if a particular characteristic is or is not a diffuse status characteristic.

The instantiation or instance of a diffuse status characteristic can be very different in different contexts. While skin colour is a diffuse status characteristic in many cultures, including the United States, it is not always a diffuse status characteristic. So, for example in South Sudan in 2016, skin colour is not a relevant status characteristic but tribal membership is.

The point is that the world changes rapidly and we can explore the changes by asking descriptive questions that provide answers about the world in a specific time. However, such description may not help us in the future, because, by definition, ordinary concepts featured in description are tied to a particular time. Theoretical principles, on the other hand, because they are not tied to a particular time, are designed to generalize and apply to the future.

Cross-cultural exploration

If experiments are designed to test theoretical principles, how are tests across cultures important? John Darley, in his Presidential Column in the newsletter of the Association for Psychological Science (2001), noted the importance of careful experimental design to enable causal inference. He stated that a basic psychological approach is to discover 'truths of human functioning that transcend culture and context.' But, 'Unfortunately, a nasty thing happened on our way to our universal generalizations: culture and context turned out to have a much more fundamental effect on our generalizations than we expected.' (Darley 2001. p. 3) Darley's statement should not be taken as an argument against the development of generalizable, theoretical principles. It can be read as an argument for more cross-cultural investigations.

If experiments are only tested in one context, for example, in a college context in predominantly White institutions in the United States, then results run the risk of support only in that particular context. At least their applicability to other settings is undemonstrated. And, indeed, psychology has been called to task by Arnett (2008) who investigated six of the APA journals and found that 'research in major APA journals is concentrated on a narrow range of the world's human population, principally Americans' (Arnett 2008, p. 609).

Replication is always important; indeed, it is one of the most important defining characteristics of science. But replication in very different contexts is especially of high value because it gives the researcher more confidence that the general principles apply even given quite different initial conditions. So cross-cultural replication is *particularly* valuable because it is a low probability event by chance alone.

Examples of such experiments include those that do replicate and do not replicate: both are important. For further discussion of this point, see Foschi 1980.

Gachter, Herrmann, and Thoni (2010) demonstrate an example of cross-cultural experiments that illustrated problems with an accepted paradigm. Basically, economic models (*Homo economicus*) suggested that cultural background should not matter because, in those models, selfishness is universal. If that is the case, then results from studies in which there is a clear prediction based upon self-interest should apply all the time, in all cultures. But, recent evidence from public goods experiments shows that is not the case.

Public goods games are a staple of economics literature; they are one of four canonical games or experimental investigations in economics (Eckel 2014). They are important to many disciplines including sociology, psychology and political science. Public goods create an individual level dilemma because individual incentives conflict with overall group incentives. While it is an actor's self-interest to *not* contribute to a good, if no one contributes, the public good will not be provided and all will be worse off. An example might be developing a public park. All could benefit and enjoy the park, independently of whether each person had contributed. But, according to the basic logic of rational choice, individuals would recognize that their own contributions are unnecessary. It is rational to not contribute, that is, to free ride. Because that is the case, traditional economic models suggested that intervention, in the sense of a central government or institution, was necessary to create and maintain public goods.

A particular *experimental* paradigm for the investigation of public goods has been developed over the years. Group members are given tokens that they can either give to the public good or keep in an individual fund. At every point in time, the value of a token kept in their fund is worth more than a token put into the public good fund. Additionally, any tokens contributed to the group fund are distributed to all members, regardless of whether or not they contributed. So, basically, the actor's best strategy is to never contribute and hope that everyone else does contribute. Of course, if everybody is thinking the same way, nobody will contribute and the public good will not be provided.

As mentioned, if the traditional economic model adequately predicted behaviour, no group member would ever contribute. Consequently, economists conjectured that there should be little if any difference across cultures because while people might make errors in judgment or perhaps be of different 'types' or personality differences tied to altruism, culture should not impact their decisions.

Herrmann, Thoni, and Gachter (2008) conducted public goods experiments in 16 different subject pools and six distinct cultural areas around the world. They did find variation however:

> Our main findings are that cooperation within cultures is largely similar while there exist highly significant differences between cultures. This is true in public good experiments with and without punishment and also holds for punishment behaviour. This dual observation of within-culture similarity and cross-cultural heterogeneity is the main support for the claim that there are cultural influences on cooperation. (Gachter, Herrmann, Thoni 2010, p. 2653).

In other words, the claim cannot be made that people respond in the same self-interested way across cultures (see also the discussion in Henrich et al. 2001).

Cross-cultural description?

As we discussed above, the strength of experiments is the testing of theoretical principles, which employ exact class concepts. Experiments are not suitable for describing a culture or the individuals who participate in experiments. Experimental data are the dependent variables of the experiment, which, as noted earlier, will be defined independently of time and place. People who participate in an experiment are not selected to represent any natural group. Rather, they are selected because it is possible to instantiate certain theoretical concepts in them. For instance, a group of young people differing on age might be selected because, for them, age meets the theoretical definition of a status characteristic. The theoretical principles under test describe how status affects behaviour. If age is a status characteristic for a particular population, then the principles can be tested by observing their behaviour. But in another time and place, age might not be a status characteristic, and then the theory would have little to say about how an age-differentiated group might behave. Experiments create artificial settings to more precisely test principles. Because they use control and random assignment, rather than trying to duplicate elements of the setting, or using random sampling, they are not useful for description.

As mentioned above, Hermann, Thoni and Gachter conducted multiple replications of public goods studies that demonstrated large amounts of cultural variation. This called into question the universality of the economic principles of self-interest in public goods settings. In 2008 and then in 2010, researchers took a different strategy and tried to interpret findings in terms of culture. They divided up the cultures into different kinds of classifications to find what factors might lead to the differences in experimental results. So, the authors suggested that 'punishment may be related to social norms of cooperation,' (Herrmann, Thoni, Gachter 2008, p. 1365). To determine this, they constructed two variables. One, norms of civic cooperation, was developed from the World Values Survey and was based upon how people feel about tax evasion, benefit fraud, and avoiding paying for public transport. A second measure was a 'rule of law' indicator based on the degree to which people abide by and believe in the rules of society including the police and the courts. Researchers reasoned that if these indicators are views of the average citizen, they also typify the participants in different cultural contexts or countries. Then they ran different analyses and found that the classification of norms of civic cooperation was related to punishment in the public goods experiments but not to the rule of law differences across cultures.

The researchers believe that the experimental results combined with the classifications tell us about the norms in different cultural settings. But this is really not clear. The questionnaire measures, if they involve random sampling, are most likely to be measures of norms for different societies. But how exactly those norms relate to behaviour in the artificial, laboratory settings is not developed. It is very unlikely that the researchers are interested in a *description* of how people in an experimental

laboratory, interacting over a computer, contribute and then levy costs (punish) others that they cannot see. Are both the questionnaire information and the experimental information tapping the same norms? If they are, then what basic principles are being tested?

In other words, experiments can be designed to see how different kinds of contexts affect the dependent variables. They can do this by utilizing control so that differences among or between conditions can be attributed to the independent variable. But when the independent variable is country, there are too many factors operating to be sure what is causing what. Experiments cannot be used to characterize a population. Populations can be characterized by methods specifically designed to measure characteristics of interest that usually employ relatively large random samples.

But how then can country or culture be incorporated into experiments? Replication is one important way. Replication would mean that the same principles are being tested through experimental designs in different cultures, not to describe the cultures, but rather to see if the same theoretical principles are upheld. To test of how framing of a social dilemma changed the degree of cooperation among group members even *when the payoffs were exactly the same*, Sell et al. (2002) conducted experiments in both the United States and China. In both countries, participants were randomly assigned to two different conditions (a 'give some' public good or a 'refrain from taking' public good) and payoffs were calibrated for each country so that they were approximately equal in meaning. Such a design is equivalent to blocking on variables for which there are good reasons to expect differences. Such blocking designs are common when investigating gender, or race/ethnicity, for example. It is obvious that we cannot assign participants an ethnicity or a country, but we can block or control on the variable. In the case of the Sell et al. (2002) experiment, participants from both countries were affected in the same way by the framing and were more cooperative in the 'refrain from taking scenario' vs. the 'give some' scenario.

Another way that cross cultural experiments can allow comparison is by ensuring that there is a type of baseline condition that measures the 'initial condition' of one culture versus another. It would not tap a genuine 'norm' in society but, importantly, it would be a measure that would allow estimation for how different theoretical changes would affect the behaviour. That is, it would serve as the beginning place for the particular experimental design. In the study of cooperation in public goods and the effects of group membership on cooperation for example, a baseline of no information about group members, would function as a 'cultural calibration' measure. This measure would then be used to determine whether and how different kinds of group membership changed initial levels of cooperation. In this way, the no information condition measures behaviour for the specific experiment, not the society as a whole. Societal norms are driven by context, and the experimental context is peculiar – it is artificial.

Oh (2013) provides an example of a cross-cultural experiment that used both initial conditions to gain an estimate of cultural differences within the experimental context, as well as a replication of theoretical (exact class) conformity principles. In this study, Oh investigated whether participants from a collectivist culture of India would conform to groups similarly to the participants from the individualist culture

of the United States. To test his conjectures, Oh designed three experiments. The first two of them were preliminary experiments designed to function as initial conditions. In these studies, Oh first tapped each of the groups' individual judgments about different choice dilemmas and different opinion items. He then used these items to generate arguments that might be posed by groups in different cultures, as well as to assess the 'cultural starting place' for this particular experimental context. As an additional baseline, he conducted a second study to determine how individual responses might change from one point in time to another for participants in each culture. The third experiment tested how participants might respond to 'mere exposure' of the opinions of groups who share their identity, or how participants respond to persuasive arguments presented by group members. Oh finds that the effects of group argumentation are the same across cultures while 'mere exposure' effects are somewhat stronger in India.

Oh's study on conformity is an excellent example of the potential strengths of cross-cultural experiments. He uses the culture as 'block variable' and he measures the initial conditions for each of the cultures, to estimate the effects of the manipulation. The manipulation is a test of the concept of conformity, a concept that is not ordinary because it does not relate to any particular context. Instead, it refers only to the change from an individuals' initial choice produced through exposure to the group.

The studies discussed illustrate that experiments are not effective at tapping descriptive properties of a particular culture; that is, they cannot adequately capture cultural norms or common characteristics of entire populations. Experiments *are* effective at testing general principles and this can be done by carefully creating baseline conditions to assess cultural 'starting places' or by replicating studies through blocking (usually by country) and random assignment.

Conclusion

Experiments are powerful methods for testing theoretical principles. They enable the manipulation of independent variables to determine the effect upon the dependent variables. They do this by creating artificial settings that enable control and randomization. Cross-cultural experiments can be important for two purposes. First, they provide important replications. Replication across cultures is especially valuable because it demonstrates that general principles apply even in very different contexts and initial conditions. Secondly, cross-cultural experiments can explore how general principles are affected by the culturally specific initial conditions.

What cross-cultural experiments cannot do effectively is describe the characteristics of cultures. Because experiments are artificial, they are not adequate methods for describing what exists in settings that are time and space specific. Experiments cannot describe the norms for littering in Poland in 2016, or the attitudes of voters in Texas in 2016. Other methods, however, can be fruitfully employed for such investigations.

References

Arnett J.J. (2008). *The Neglected 95%: Why American Psychology Needs to Become Less American.* American Psychologist, 63, p. 602–614.

Cohen B.P. (2003). *Creating, Testing, and Applying Social Psychological Theories.* Social Psychology Quarterly, 66, p. 5–16.

Cohen B.P. (1989). *Developing Sociological Knowledge: Theory and Method*, 2nd ed., Nelson-Hall, Chicago, IL.

Darley J. (2001). *We Fail to Contribute to Policy Debates.* APS Observer, 14, p. 3–40.

Eckel C. (2014). *Economic Games for Social Scientists*, In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. London: Elsevier, p. 335–356.

Foschi M. (1997). *On Scope Conditions.* Small Group Research, 28, p. 535–555.

Foschi M. (1980). *Theory, Experimentation, and Cross-Cultural Comparisons in Social Psychology.* Canadian Journal of Sociology, 5, p. 91–102.

Gachter S., Herrmann B., Thoni C. (2010). *Culture and Cooperation.* Philosophical Transactions of the Royal Society of London B: Biological Sciences, 365, p. 2651–2661.

Henrich J., Boyd R.,. Bowles S., Camerer C., Fehr E., Gintis H., McElreath R. (2001). *In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies.* American Economic Review, 91, p. 73–78.

Herrmann B., Thoni C., Gachter S. (2008). *Antisocial Punishment Across Societies.* Science, 319, p. 1362–1367.

Korner S. (1966). *Experience and Theory: An Essay in the Philosophy of Science.* London: Routledge and Kegan Paul.

Oh S.E. (2013). *Do Collectivists Conform More than Individualists? Cross-Cultural Differences in Compliance and Internalization.* Social Behavior and Personality, 41, p. 981–994.

Ridgeway C.L., Backor K., Li Y.E., Tinkler J.E., Erickson K.G. (2009). *How Easily Does a Social Difference Become a Status Distinction? Gender Matters.* American Sociological Review, 74, p. 44–62.

Ridgeway C.L., Boyle E.H., Kuipers K., Robinson D.T. (1998). *How Do Status Beliefs Develop? The Role of Resources and Interactional Experience.* American Sociological Review, 63, p. 331–350.

Ridgeway C.L., Correll S.J. (2006). *Consensus and the Creation of Status Beliefs.* Social Forces, 85, p. 431–454.

Ridgeway C.L., Erickson K.G. (2000*). Creating and Spreading Status Beliefs.* American Journal of Sociology, 106, p. 579–615.

Sell J., Chen Z.Y., Hunter-Holmes P., Johansson A. (2002). *Cross-cultural Comparison of Resource Goods and Public Goods.* Social Psychology Quarterly, 65, p. 285–297.

Sell J., Martin M.W. (1983). *An Acultural Perspective on Experimental Social Psychology.* Personality and Social Psychology Bulletin, 9, p. 345–350.

Walker H.A., Cohen B.P. (1985). *Scope Statements: Imperatives for Evaluating Theory.* American Sociological Review, 50, p. 288–301.

Webster M., Jr., Hysom S.J. (1998). *Creating Status Characteristics.* American Sociological Review, 63, p. 351–378.

Webster M. Jr., Sell J. (2014). *Why do experiments?* In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. New York: Elsevier, p. 5–22.

Znaczenie eksperymentów międzykulturowych dla nauk społecznych

Wyszczególniając czym eksperyment różni się od innych typów badań, wskazujemy na znaczenie kontroli i randomizacji w eksperymentach laboratoryjnych oraz potrzebę tworzenia sztucznych środowisk, w których możliwe staje się sprawdzanie szczególnego rodzaju teorii – teorii zbudowanych przy użyciu pojęć nie ograniczonych czasem i przestrzenią. Celem eksperymentów międzykulturowych jest zapewnienie, by testowanie teorii nie było przeprowadzane tylko w jednym układzie kulturowym. Takie eksperymenty mogą także ułatwić poznanie swoistych dla różnych kultur warunków początkowych (szczególnych realizacji warunków zakresowych), co służy dalszemu rozwijaniu teorii.

Słowa kluczowe: eksperymenty laboratoryjne, badania międzykulturowe, randomizacja, warunki początkowe