EXPERIMENTATION IN PSYCHOLOGY—RATIONALE, CONCEPTS, AND ISSUES

Siu L. Chow
Department of Psychology, University of Regina, Canada

Keywords: conditional syllogism, control, criterion of falsification, experiment, falsification, generality, hypothesis, induction, modus ponens, modus tollens, social psychology of the psychological experiment, theoretical prescription, validity, variable, verification

Contents

Summary

The experiment is an arrangement for collecting research data, in which there are two or more conditions that are identical in all aspects but one. The aspect in which the test

conditions differ is the independent variable. Both deductive and inductive logic are used in experimentation, albeit at different stages for different purposes. While deductive logic is used to derive the experimental hypothesis from the substantive hypothesis, inductive logic is the foundation of the experimental design. The theoretically informed control variables and methodologically informed control procedures are responsible for the feature that differentiates the experiment from non-experimental studies. The feature in question is the provision for experimental controls, the function of which is to exclude recognized alternative explanations. The three control features are (a) a valid comparison baseline, (b) the constancy of conditions, and (c) provisions for excluding procedural artifacts.

Given the differences in impetus and objectives, utilitarian and theory-corroboration experiments differ also in their theoretical foundation and their proximity to real-life phenomena. Much of the misunderstanding of, as well as the dissatisfaction with, the experimental approach is because theory-corroboration experiments are discussed and assessed with criteria that are more appropriate for utilitarian experiments. For example, it is not readily seen from the utilitarian experiment that (a) experimental data owe their meaning to three embedding conditional syllogisms, and (b) ecological validity is irrelevant, if not harmful altogether.

The experimental approach to theory corroboration can be defended in the present relativistic milieu because of its control provisions. Moreover, it has been shown that there are no grounds for the critique in terms of the social psychology of the experiment.

1. Introduction

Skinner once said that conducting experiments involved nothing more than measuring subjects' simple, countable behaviors while manipulating some randomly selected aspect of the environment. No planning is required. Skinner gave the impression that experimentation is a chancy trial-and-error exercise suitable only for studying simple, countable phenomena that can be shaped by experimenters. However, Skinner's conclusions about operant conditioning were actually based on carefully designed experiments that satisfy sophisticated inductive principles. The "trial-and-error" nature of experimentation is actually a characteristic of the Popperian "conjectures and refutations" endeavor at the conceptual level.

Empirical research must have at least two test conditions that satisfy certain stipulations before it can be characterized as an experiment. The present discussion begins with a description of the components of the experiment. A distinction is then made between utilitarian and theory-corroboration experiments. The relationship between the experimental and control conditions is then explained by making explicit the roles of deductive and inductive logic at various stages of experimentation. The discussion concludes with some meta-theoretical issues and their implications.

2. Components of the Experiment

Consider an experiment conducted to investigate the effects of music on mood. Subjects adapt to a piece of music in either the major or minor key in Phase I. Of experimental interest is the subjects' recognition performance on Task T while being exposed to the same piece of music in Phase II. The experiment has five components: (a) Task T in Phase II, (b) the procedure (which includes the adaptation in Phase I), (c) its three explicit types of variables, (d) its design, and (e) the inductive principle that underlies its design. These features may be introduced with reference to Table 1.

		Independent variable	**Control variables**				**Dependent variable**
			AD[a]	**Tempo**	**Timbre**	**Performer**	
1	Experimental	Major key	5 secs	Mod[b]	Piano	Joe	$d' = 0.82$[c]
2	Control	Minor key	5 secs	Mod	Piano	Joe	$d' = 0.67$

[a] AD = adaptation duration = the duration of the adaptation in Phase I
[b] Mod = moderate tempo
[c] $d' = 0.82$: this is an index of sensitivity, a larger number means a greater sensitivity

Table 1. The design of the 1-factor, 2-level experiment used to assess the effects of the musical key on recognition performance

2.1. Types of Variables

A variable is anything that can be identified in more than one way. For example, *musical key* in Table 1 is a variable because it is represented either by a major key or a minor key. In conducting an experiment, psychologists manipulate the independent variable and measure the dependent variable while holding the control variables constant. There are also the extraneous variables that, while not explicitly identified, are nonetheless assumed to have been held constant by virtue of the appropriate control procedures found in the experiment.

2.1.1. The Independent, Control, and Dependent Variables

The psychologist in Table 1 manipulates musical key by setting up two test conditions, one with a piece of music in a major key and another with a piece of music in a minor key. Musical key is the independent variable in the sense that the two conditions are set up independently of what the subjects do. As may be seen from Table 1, *adaptation duration* (i.e. the time spent listening to music in Phase I), *tempo*, *timbre* and *performer* are held constant when the experimenter uses the same level of each of them in both the major-key and minor-key conditions. In such a capacity, they are the control variables. Subjects' recognition performance is measured (e.g. with the index of sensitivity d'). It is the dependent variable because its values depend on the subjects.

2.1.2. The Extraneous Variable, Confounding Variable, and Control Procedure

Any variable that is not the independent or the dependent or the control variable is an extraneous variable. Although there are logically an infinite number of extraneous variables, it is possible to eliminate most of them on conceptual or theoretical grounds. For example, it is reasonable to exclude variables such as the preference for cereal and height as possible explanations of the data in the example in Table 1.

Given the design depicted in Table 1, the loudness of the music is an extraneous variable assumed to be irrelevant or to have been held constant. Suppose that it is discovered at the conclusion of the experiment that the major-key music is louder than the minor-key music. Varying systematically with musical key, *loudness* becomes a confounding variable. It renders ambiguous the meaning of the data (for something similar in quasi-experiments **see *Quasi-Experimentation***).

2.2. Experimental Designs

The next feature of the experiment is its design: the formal arrangement of (a) the independent, dependent, and control variables, (b) the sequence of events in the course of the experiment, (c) the sequence of events to be carried out in a trial, and (d) how subjects are assigned to the test conditions. The numerous experimental designs may be categorized in terms of (i) the number of independent or dependent variables, and (ii) the manner in which the subjects are assigned to the test conditions.

2.2.1. Designs and the Number of Variables

A distinction is made between univariate and multivariate designs. There is only one dependent variable in the univariate design, whereas two or more dependent variables are used in the multivariate design. Experimental designs are also classified in terms of the number of independent variables: 1-factor design (i.e. one independent variable) and multifactor design (two or more independent variables). Regardless of the number of independent variables used, designs are further identified in terms of the number of levels used to represent the independent variables. For example, designs involving only one independent variable may further be distinguished between the "1-factor, 2-level" and "1-factor, multilevel" varieties. As the names suggest, only two levels of the independent variable are used in the former, and more than two levels are used in the latter.

Multifactor designs may be complete factorial or incomplete designs. Shown in Table 2 is an example of the "2 × 3 factorial" variety. The number of numerals in the name indicates the number of independent variables, and the identity of each of the numerals represents the number of levels used to represent that independent variable. Hence, the two numerals in "2 × 3" means that there are two independent variables. The first variable has two levels whilst the second variable has three levels. This convention makes it easy to describe any design. As another example, the 3 × 2 × 4 × 5 design means that there are four independent variables. They have three, two, four, and five levels, respectively.

		Variable B		
		$\mathbf{B_1}$	$\mathbf{B_2}$	$\mathbf{B_3}$
Variable	A_1	AB_{11}	AB_{12}	AB_{13}
A	A_2	AB_{21}	AB_{22}	AB_{23}

Table 2. The schematic representation of the 2 × 3 design

Represented in the two rows of Table 2 are the two levels of A (A_1 and A_2). The three levels of B are represented in three columns (B_1, B_2, and B_3). Variables A and B jointly define six treatment combinations: AB_{11}, AB_{12}, AB_{13}, AB_{21}, AB_{22}, and AB_{23}. In other words, a treatment combination is a specific test condition that is defined by a combination of specific levels of two or more independent variables.

There are six treatment combinations in the 2×3 design. In fact, the total number of treatment combinations in a multifactor experiment is the product of the respective numbers of levels of the independent variables found in the design. For example, there are 120 treatment combinations in the $3 \times 2 \times 4 \times 5$ factorial design.

The distinction between a complete factorial and incomplete design may now be explained. A complete factorial design is one in which data are available from every treatment combination. It becomes an incomplete design if data are missing from one or more of the treatment combinations. Researchers avoid using the incomplete design as much as possible because both data analysis and data interpretation are difficult when incomplete designs are used.

2.2.2. Designs and Subject Assignment

In terms of how subjects are assigned to the test conditions, the 1-factor design may be a completely randomized or a repeated-measures design. When the completely randomized design is used, subjects are assigned randomly to the test conditions. By "random assignment" is meant that whoever is included in one condition does not determine, or is not determined by, whoever is assigned to another condition. In contrast, the same subject is tested in every test condition found in the experiment when the repeated-measures design is used.

In the same vein, multifactor factorial designs fall into four categories: the completely randomized, the repeated-measures, the randomized block, and the split-plot designs. Recall that there are numerous extraneous variables in any experiment. Suppose that individuals are not assigned randomly to the test conditions. Instead, those who have won a scholarship are assigned to the major-key condition, whereas individuals who have just failed a quiz are assigned to the minor-key condition. Common sense suggests that the major-key group is a happy group and the minor-key group is a less happy group to begin with. Under such circumstances, *being happy* is a confounding variable because its two levels are yoked to those of the independent variable. Hence, any difference between the two musical-key conditions could have been due to the differences between the two levels of being happy.

In short, no extraneous variable should be confounded with the independent variable. The sole purpose of using the completely randomized design is to minimize such confounding. Random subject assignment ensures that, in the long run, the ratio of being-happy to being-less-happy subjects would be the same at both levels of musical key.

In the event that random subject assignment is insufficient for holding constant a potential confounding variable, experimenters may test every subject in all test conditions (i.e. using the repeated-measures design). This procedure may minimize confounding. Consider the experiment described in Table 1. Regardless of individual "happiness" level, it would be the same in the major-key and minor-key conditions if an individual is being tested in both of them.

However, it is not always possible to use the repeated-measures option. First, subject fatigue may become an issue when they are being tested for a longer period of time. Second, there is also the potential difficulty due to the order of testing, as may be seen from Table 3. The order of testing in row 1 is AB_{11}, AB_{12}, AB_{21}, and AB_{22}. The possible source of ambiguity is that the subjects' performance in any of the other three treatment combinations might be different had they not been tested previously in AB_{11}. The same difficulty applies when subjects are tested first in AB_{12}, AB_{21}, or AB_{22}. This source of ambiguity is one exemplification of the "order of testing" effects.

	Order of testing			
	1	**2**	**3**	**4**
Group 1	AB_{11}	AB_{12}	AB_{21}	AB_{22}
Group 2	AB_{12}	AB_{21}	AB_{22}	AB_{11}
Group 3	AB_{21}	AB_{22}	AB_{11}	AB_{12}
Group 4	AB_{22}	AB_{11}	AB_{12}	AB_{21}

Table 3. The schematic representation of a Latin-square arrangement

A solution to this particular form of order of testing effect is to divide the subjects into as many groups as there are treatment combinations (four in the present example). The groups are tested in the particular orders that make up a Latin square, as shown in Table 3. The four groups collectively ensure that any treatment combination is tested at each of the four temporal positions equally often. Consequently, this Latin-square arrangement ensures that data are balanced in terms of "being tested first." However, this is only a partial solution to the order of testing difficulty because the arrangement in Table 3 leaves many testing orders unbalanced.

There is a more serious constraint on the applicability of the repeated-measures design. Some experimental manipulations produce an irrevocable result, for example, *therapeutic method*. Its two levels may be surgery and medication. Having undergone surgery, the subject cannot be restored to the pre-surgery state in order to be tested under the medication condition. It is in this context that the randomized block design depicted in Table 4 may be appreciated.

	Treatment combination			
Groups matched in terms of musical sophistication	AB_{11}	AB_{12}	AB_{21}	AB_{22}
Expert (E)	E_2	E_4	E_1	E_3
Moderate (M)	M_2	M_1	M_4	M_3
Novice (N)	N_4	N_1	N_2	N_3

The subscripts of AB represent the treatment-combination.
The subject E, M, or N refers to an individual in the block.

Table 4. The schematic representation of the randomized block design

Prospective subjects are first pre-tested for their musical sophistication, and put in one of three groups (expert, moderate, or novice). The procedure ensures that the within-group homogeneity (in musical sophistication) is higher than between-group homogeneity. The size of the groups is some multiple of the number of test conditions. This stipulation ensures that all treatment combinations have the same number of subjects from every level of musical sophistication.

The fourth type of design in terms of subject assignment is the split-plot design when there are two or more independent variables. Given the 2×3 factorial design in Table 2, it is possible to use Variable A as the repeated-measures variable (i.e. the same subjects are tested at both levels of A), but different subjects are randomly assigned to the three levels of B. Such an arrangement is an example of the split-plot design.

2.3. The Inductive Foundation

The experimental design owes its importance to its underlying inductive principle whose function is to reduce ambiguity in data interpretation. As an illustration, underlying the 1-factor, 2-level design depicted in Table 1 is J.S. Mill's method of difference. The idea is to set up two conditions that are identical in all aspects except one. Specifically, the two levels of the independent variable are used to set up the two otherwise identical conditions.

The force of the method of difference (or any inductive principle) is a negative one (albeit very important). Note specifically that a difference in *d*' is found despite the fact that *adaptation duration*, *tempo*, *timbre*, and *performer* are held constant in both conditions. In other words, they are irrelevant to the observed difference in *d*'. Consequently, they can be excluded as explanations of the data. That is, the inductive principle makes it possible to exclude specific alternative causes.

2.4. Three Technical Meanings of "Control"

The important interpretation-exclusion function of inductive logic is encapsulated in the experimental control whose three components are (i) the provision for excluding confounding variables, (ii) the constancy of condition, and (iii) the valid comparison baseline.

First, the possibility of having a confounding variable may be minimized by a procedure such as randomizing the order of stimulus presentation or counterbalancing the order of testing if the repeated-measures design is used. Either of them is a control procedure used to exclude a possible artifact. Although control procedures cannot be seen from the schematic representation of design (such as Table 1), they are (or should be) described in full in the "Procedure" section of the experimental report.

Second, there are two aspects to the constancy of condition in the experiment. The first is the stipulation that the predetermined levels of the independent variable be applied consistently throughout the experiment. For example, if the two pieces of music used in the experiment begin in the key of C major and C minor, these should be used throughout the experiment. The second aspect is the better-known provision of control variables, as illustrated by the variables *adaptation duration*, *tempo*, *timbre*, and *performer* in Table 1. They are held constant in the sense that the same level of each is used at the two levels of musical key (for the attempts to achieve constancy of condition in non-experimental research, **see *Interviewing and Observation***).

Third, to be able to conclude that the difference between the two musical-key conditions is not due to an artifact, it is necessary to ensure that the test conditions are identical in all aspects except for the level of musical key. If the repeated-measures design is used, either of the two musical key conditions in Table 1 satisfies this stipulation when it is used to assess the effect of the other level.

In sum, the three components of experimental control serve collectively to exclude explanations other than the independent variable. This is very different from the Skinnerian use of "control" because, as has been shown, experimental control has nothing to do with constraining or shaping what experimental subjects do. If one were to use "control" in the Skinnerian sense to mean constraining or shaping behavior, it is researchers' data-interpretation that is being constrained or shaped. Specifically, researchers are prohibited logically from appealing to factors that are used explicitly as control variables or procedures.

-
-
-

Bibliography

Cohen M.R. and Nagel E. (1934). *An Introduction to Logic and Scientific Method*, 467 pp. London: Routledge and Kegan Paul. [These authors showed why Mill's Method of Agreement was unsatisfactory and that Mill's other canons of inductive reasoning had an important negative function, namely, as a means to exclude alternative explanations.]

Hull C.L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*, 422 pp. New York: Appleton-Century. [Many criticisms of the hypothetico-deductive method to theory corroboration are based on Hull's approach. However, those criticisms do not apply when the hypothetico-deductive method is used from the Popperian perspective.]

Mill J.S. (1973). *A System of Logic: Rationcinative and Inductive*. Toronto: University of Toronto Press. [This is the primary source about inductive principles that are more sophisticated than the commonly assumed induction by enumeration. However, Mill's view that inductive principles can be used to identify causes is problematic.]

Skinner B.F. (1938). *The Behavior of Organisms: An Experimental Analysis*, 457 pp. New York: Appleton-Century. [This book is a good illustration of the fact that a good experimenter is not necessarily

a good commentator on experimentation as a methodology. Despite his claims, his data are accepted because they come from properly designed experiments.]

Biographical Sketch

Siu L. Chow obtained his B.A. Honours from the University of Adelaide, Australia, and his Ph.D. from the University of Toronto, Canada. He is currently professor of psychology at the University of Regina, Canada.