Chow, S. L. (2002). EXPERIMENTATION IN PSYCHOLOGY—RATIONALE, CONCEPTS, AND ISSUES. In *Methods in Psychological Research*, In Encyclopedia of Life Support Systems (EOLSS), Eolss Publishers, Oxford, UK. [http://www.eolss.net]

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.

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

		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

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.

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

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

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.

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

	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}

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.

Table 4. The schematic representation of the randomized block design

	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.

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.

3. Types of Experiments

Although all experiments share the same formal structure, two types of experiments may be distinguished in terms of the set of important issues listed in column 1 of Table 5. Experiments in the utilitarian tradition are atheoretical (i.e. indifferent to explanatory theories), the impetus of which is a practical question. The aim is to ascertain the causal efficacy of some substantive agent (e.g. a new drug) or procedure (e.g. using the phonic method to teach beginning readers). Hence, column 2 is the "Utilitarian" column (for the theoretical versus atheoretical distinction in non-experimental research, see *Interviewing and Observation*).

In contrast, regardless of their practical relevancy, many psychological phenomena invite explanations. For example, most drivers have the experience of realizing belatedly that they have gone through a red light without stopping. Why do drivers "see" the red light belatedly? The tenability of an explanation for phenomena like this has to be substantiated empirically. This requires a different sort of experiment: the theory-corroboration experiment.

Table 5. Some differences between experiments used to establish a functional relationship and theory-corroboration experiments

	1	**2 Utilitarian**	**3 Theory-corroboration**
1	Impetus	A practical question—Does the phonic method improve reading skills?	A phenomenon that invites an explanation—Why can we "see" the red light belatedly?
2	Purpose	To ascertain the efficacy of an agent or procedure	To test an implication of an explanatory theory
3	Theory	Non-explanatory functional relationship	Explanatory theory with hypothetical mechanisms
4	Cause	Efficient cause	Material or formal cause
5	Experimental hypothesis	Some re-phrasing of the practical question	An implication of the theory in a particular experimental context
6	Independent variable	The substantive agent or procedure itself	A theoretically informed manipulation that is not predicated on the phenomenon
7	Dependent variable	The result of applying the to-be-assessed agent or procedure	Different from the to-be-explained phenomenon
8	Intervening variables	Logical constructs	Hypothetical structure and mechanisms imbued with testable theoretical properties
9	Control variables	No need to change in replications	Theoretically informed
10	Explanation	A demonstration of class membership (being an example of the functional relationship)	What is said about the properties of the hypothetical entity in the explanatory theory
11	Replication	No change in the experimental hypothesis	A different implication of the explanatory theory
12	Use of the findings	To give direction to a practical course of action—prediction and control of behavior	To assess the tenability of the explanatory theory
13	Effect size	May be important in the presence of an additional criterion	Irrelevant
14	Ecological validity	Necessary	Irrelevant if not detrimental

The utilitarian objective of establishing functional relationships between observable variables is to be justified by demonstrating regularities (e.g. manipulating X produces Y regularly). Such a reliable functional relationship renders it possible to shape behavior (Y) by manipulating X (hence, the "Efficient cause" entry in row 4 in Table 5). On the other hand, explanatory theories have to be empirically substantiated because they implicate unobservable hypothetical structures or mechanisms that are imbued with theoretical properties. Theory-corroboration experimenters have no illusion of shaping behavior. Instead, they are concerned with the material or formal cause of the phenomenon (that we behave the way we do because our psychological apparatus has certain properties or structure). The meta-theoretical differences between the two approaches to psychological phenomena have important implications for experimentation (rows 5 through 14 in Table 5).

4. The Utilitarian Experiment

The experimental hypothesis of a utilitarian experiment is a simple re-phrasing of the practical question itself. For example, if the concern is about the relative merits of two methods of teaching beginning readers, the experimental question is simply “Do the phonic and whole-word methods produce different reading skills?” or “Does the phonic method produce higher reading performance than the whole-word method?” It follows that *teaching method* has to be the independent variable, and its experimental level has to be the phonic method of teaching. Moreover, the dependent variable has to be the reading performance of beginning readers.

Experimenters in the utilitarian tradition do not often find it necessary to examine the reason why functional relations are what they are. If the “why” question is ever asked, the explanation would be an appeal to the regularity demonstrated. Nonetheless, they may invoke intervening variables as logical constructs so as to economize descriptive efforts. Literal replications are essential for the regularity envisaged. Hence, nothing is (or can be) changed in attempts to replicate experiments of this genre.

Given the utilitarian objective, findings from these experiments are used to guide a practical course of action. For this reason, ecological validity is deemed important (i.e. the test situation and the experimental task must be as similar as possible to their counterparts found in the real-life phenomenon). Moreover, the magnitude of the substantive manipulation’s efficacy is important. However, this magnitude is not the effect size dealt with in statistics (see *Statistics and its Role in Psychological Research*).

5. The Theory-Corroboration Experiment

The situation is very different with theory-corroboration experiments. The experimental question is not (and cannot be) the research impetus. To appreciate fully the logical foundation of theory-corroboration experimentation, it is necessary to distinguish between the explanatory theory (or substantive hypothesis), the research hypothesis, the experimental hypothesis, and the statistical hypothesis.

5.1. Phenomenon and Its Explanation

Suppose that the phenomenon of interest is “seeing” belatedly a red light. Relevant to the understanding of the phenomenon is the fact that, paying attention to something else while driving, drivers fail to identify the red light ahead. To make sense of the phenomenon, it may be speculated that, so long as the drivers’ eyes are oriented in the correct direction, the red light is registered in the sensory visual store automatically. The said sensory information remains in the large-capacity, veridical, visual sensory store if it is not obstructed or erased by other visual inputs, and can be processed within 0.5 seconds. To the extent that there is empirical evidence in support of the speculation, the phenomenon is explained.

As may be seen from Table 6, the research question is not about the phenomenon itself, but about its explanation (for a similar distinction in psychometrics, see *The Construction and Use of Psychological Tests and Measures*). Nor can the phenomenon itself be used as the evidential support for the theory. To do so would be putting forward a circular argument. At the same time, it is not possible to answer the question “Is there a transient, large-capacity, veridical visual buffer?” in the way “Is there a chair in front of us?” is answered for the simple reason that the said visual buffer is unobservable.

Table 6. Distinguishing between the substantive theory, the research hypothesis, the experimental hypothesis, and the statistical null hypothesis

1	Phenomenon	"Seeing" the red light belatedly
2	Explanatory theory (substantive hypothesis)	Information about the red light is available for a brief period of time by virtue of a transient, large-capacity visual buffer
3	Research hypothesis	Is there empirical evidence in support of the transient, large-capacity visual buffer?
4	Experimental hypotheses	Hypotheses implied by the theory that there is a transient, large-capacity visual buffer imbued with a set of well-defined properties that can explain the phenomenon: 1. If the large-capacity assumption is true, then there is partial-report superiority 2. If the representation is very brief, then partial-report superiority diminishes within 0.5 seconds 3. If the information is unprocessed, then there is no partial-report superiority with a category cue, but there is partial-report superiority with a spatial or colour or size cue
5	Statistical null hypotheses (H_0)	(1) H_0: $u_{\text{partial report}} = u_{\text{whole report}}$ (2) H_0: $u_{(\text{partial report} - \text{whole report})\text{at T1}} = u_{(\text{partial report} - \text{whole report})\text{at T2}}$ (3) H_0: $u_{\text{partial report}} = u_{\text{whole report}}$ regardless of cue-type

Be that as it may, for the theory to be taken seriously, the transient visual buffer envisaged must have observable consequences in some well-defined context. Part and parcel of designing the experiment is setting up a well-defined situation for data collection (for the whole-report and partial-report tasks required in dealing with the example depicted in Tables 6, 7, and 8, see Appendix 1). This makes it possible to set up the criterion for rejecting the theory.

5.2. The Criterion of Falsification

Three of numerous experimental hypotheses have been singled out in panel 4 of Table 6. They are in the form of a conditional proposition whose antecedent is a theoretical property of the postulated visual buffer. The consequent of the conditional proposition is an implication of the said property in a particular situation. Four observations may be made about the experimental hypotheses.

First, none of the experimental hypotheses is about the to-be-explained phenomenon. Second, they differ from the research hypothesis (see panel 3 of Table 6). Third, they differ among themselves in that they deal with different theoretical properties of the hypothetical visual buffer. Fourth, the consequent of the conditional proposition is a criterion of rejection of the theoretical property depicted in the antecedent (hence, a criterion of falsification for the explanatory theory). That is, the explanatory theory has to be rejected if there is no evidential support for any of its theoretical properties.

5.3. The Statistical Null Hypothesis (H_0)

Also depicted in Table 6 are the three statistical null hypotheses (H_0) for their respective experimental hypotheses. It may be seen that each H_0 is a re-formulation of the consequent of an experimental hypothesis in terms of the parameters of the statistical populations defined by the test conditions (see *Statistics and its Role in Psychological Research*). That is, H_0 is not the substantive or the research or the experimental hypothesis.

5.4. The Rationale of Theory Corroboration

It is unfortunate for conceptual clarity that positivistic terminology is still being used in methodological discussions. Specifically, it is incorrect to describe theory corroboration as an attempt to infer a theory from data. In actual fact, experimenters are assessing the preexisting theory with experimental data when the data are collected in a way informed by the to-be-corroborated theory. It is also unfortunate that the theory-corroboration process is often not distinguished from the statistical hypothesis testing procedure. A discussion of the rationale of theory-corroboration experimentation with reference to Tables 7 and 8 may help to clarify the issues.

Table 7. The hypotheses implicated in testing Experimental Hypothesis 1 identified in Table 6

1	Level of abstraction	What is found or said at the level of abstraction
2	Phenomenon	Drivers may belatedly realize that they have missed the red light
3	Theory (substantive hypothesis)	[P7.1]: A transient, large-capacity visual buffer that holds only raw sensory information makes it possible for our visual system to register more information than can be processed or recalled immediately
4	Research hypothesis	[P7.2]: **If** [P7.1] is true, **then** experimental data should assume particular patterns (Y) in certain situations (X)
5	*Complement of the research hypothesis*	*[P7.2']: **If** [7.1] is false, **then** experimental data should not be in the form of Y when being collected in X*
6	Experimental Hypothesis 1	[P7.3]: **If** only raw sensory information is available in the buffer, **then** partial-report superiority is found when a spatial, but not a category, cue is used.
7	*Complement of Experimental Hypothesis 1*	*[P7.3']: If the storage format said in [P7.1] is false, then there is no partial-report superiority with a spatial cue*
8	Statistical alternative hypothesis (H_1)	[P7.4.]: **If** [P7.3] is true, **then** H_1: $u_{\text{partial report}} > u_{\text{whole report}}$
9	*Statistical null hypothesis (H_0)*	*[P7.4']: If [P7.3] is false, then H_0: $u_{partial\ report} \leq u_{whole\ report}$.*
10	Sampling distribution based on H_1	[P7.5]: If H_1 is used, the probability associated with a *t*-value as extreme as 1.645 is **not known**
11	*Sampling distribution based on H_0*	*[P7.5']: If H_0 is used, the probability associated with a t-value as extreme as 1.645 is 0.05 in the long run*

The corroboration of Experimental Hypothesis 1 (row 6 of Table 7) involves more than the hypothesis itself. It may be seen that [P7.3] in row 6 owes its origin to [P7.1] (row 3), which implies [P7.2] (row 4). The logical complement of [P7.3] (i.e. *[P7.3']* in row 7) is re-stated as the statistical null hypothesis *[P7.4']* (row 9) on the basis of which the *t*-distribution is chosen to test H_0 (see *Statistics and its Role in Psychological Research*). The implicative relationships between [P7.1], [P7.2], [P7.3], and *[P7.4']* (i.e. rows 4, 5, 6, and 8, respectively), as well as the inferential process implicated in theory corroboration, may best be seen from Table 8.

Table 8. The embedding conditional syllogisms underlying theory corroboration when H_0 is false

1	**Major premise 3**	**If the theory of the visual buffer is true, then the visual system registers more raw visual information than can be recalled immediately**
2	*Major premise 2*	*If the visual system registers more raw visual information than can be recalled immediately, then there is partial-report superiority when a spatial cue is used*
3	Major premise 1	If there is partial-report superiority, then $u_{\text{partial report}} > u_{\text{whole report}}$
4	Minor premise 1	Data lead to the rejection of $u_{\text{partial report}} \leq u_{\text{whole report}}$
5	Conclusion 1	There is partial-report superiority in the interim
6	*Minor premise 2*	*There is partial-report superiority when a spatial cue is used*
7	*Conclusion 2*	*The visual system registers more raw visual information than can be recalled immediately in the interim*
8	**Minor premise 3**	**The visual system registers more raw visual information than can be recalled immediately**
9	**Conclusion 3**	**The visual buffer envisaged is tenable in the interim**

The innermost conditional syllogism may be found in rows 3, 4, and 5 of Table 8. The next conditional syllogism consists of the *italicized* entries in rows 2, 6, and 7. The **boldface** entries in rows 1, 8, and 9 are the major premise, minor premise, and conclusion, respectively, of the outermost conditional syllogism. The whole chain of inferences begins with "Minor premise 1" in row 8, which is supplied by the experiment.

Note that the consequent of "Major premise 1" is that there is partial-report superiority (i.e. $u_{\text{partial report}} > u_{\text{whole report}}$). Suppose that the result is not statistically significant. The experimental conclusion is that the data are inconsistent with partial-report superiority (i.e. there is no partial-report superiority). The antecedent of "Major premise 1" is then rejected by virtue of modus tollens. This in turn warrants the rejection of the antecedents of "*Major premise 2*" and **"Major premise 3"** in succession. The theoretical conclusion is that there is no empirical support for the veridical visual buffer (see *Methods in Psychological Research)*.

This does not mean that the result of a single experiment is sufficient to reject a theory. As may be seen from Appendix 1, in addition to appealing to statistical significance, it is also necessary to ensure that there is no procedural artifact. For example, data contrary to the visual buffer may be found when subjects are not given sufficient training on the partial-report task. In other words, it is misleading to say that a theory has to be rejected simply because there is no statistical significance. Researchers must be sure that the experiments that produce the statistically non-significant result have been conducted properly.

5.5. Tentative Conclusion and Converging Operations

Depicted in Table 8 is the more complicated situation in which the statistically significant experimental data make it possible to reject H_0. That is, the data support the experimental expectation that there is partial-report superiority. This amounts to affirming the consequent of "Major premise 1." The antecedent of "Major premise 1" thus affirmed is used in the interim as the minor premise of the second syllogism, in which the consequent of "*Major premise 2*" is affirmed. The antecedent of *"Major premise 2"* is then used as the minor premise of the last syllogism. It affirms the consequent of **"Major premise 3,"** and the conclusion is to affirm the antecedent of **"Major premise 3"** in the interim. However, this chain of inferences is logically problematic because affirming the consequent of a conditional proposition does not warrant affirming its antecedent. In more concrete terms, the statistical significant result in the example may be due to factors other than the putative transient visual buffer. This is why the control variables and procedures are important. They are used to exclude all *recognized* alternative explanations. That is, the exclusion function of inductive logic is used to warrant the tentative acceptance of the theory even though deductive logic does not permit such a conclusion (see *Methods in Psychological Research)*.

Nonetheless, the necessity of using the "recognized" caveat is recognition that some yet-unrecognized factors may be the actual reason for the statistical significance. Hence, the "in the interim" qualification is necessary in "Conclusion 1," *"Conclusion 2"* and **"Conclusion 3."** This is also the reason why it is not possible to prove any theory.

Be that as it may, conducting additional experiments may strengthen the tentative nature of the conclusion from a single experiment. Unlike the utilitarian approach, however, these additional experiments are not (and should not be) literal replications. Nor are replications helpful when there may be a yet-to-be-discovered design flaw in the original study. By the very nature of a replication, the design flaw is inevitably reinstated every time the same experiment is conducted.

In other words, the additional experiments in question must be new experiments testing the same implication in different situations or entirely new implications (e.g. Experimental Hypotheses 2 and 3 in panel 4 of Table 6). In fact, there are logically numerous experimental hypotheses. In testing each of them, psychologists are attempting to falsify the theory. The more often psychologists fail in their attempts to falsify the theory, the more the attempts converge on the tenability of the theory. Hence, the series of experiments used to test Experimental Hypotheses 1 through 3 constitute the converging operations carried out to corroborate the explanatory theory. Our faith in retaining the theory increases with more experiments included in the converging operations.

In sum, owing to the nature of deductive logic, there is asymmetry between the logical certainty in rejecting the theory and the tentative nature of retaining the theory. The inferential process is made up of three embedding conditional syllogisms that owe their origin to the theory. It is possible to interpret the data only because there exists some explanatory theory before data collection. Experimenters do not infer the theory from the data. Nor are the experimental data used to explain the phenomenon. The to-be-tested theory (which gives rise to the experimental hypothesis) becomes the valid explanans if it is supported by experimental data (see *Methods in Psychological Research)*.

6. Criticisms of Experimental Psychology Revisited

The experimental approach to psychological research is not without its critics. Hence, it is important to deal with the following issues: (a) How can one study objectively things that are not observable? (b) How objective are experimental data when they rely on unreliable observations? (c) How is objectivity possible if observations are theory dependent? (d) Conducting experiments implicates social interactions between experimenters and subjects. How is objectivity possible in view of the social psychology of the psychological experiment?

6.1. Empirical Study of the Unobservable

The feasibility of studying the unobservable has been anticipated and dealt with in *Section 5.1. Phenomenon and Its Explanation*. That is, so long as the unobservable hypothetical structure is well defined enough to prescribe an observable consequence in a specific context, it is possible to design and conduct the appropriate experiment to test the hypothetical structure (see Tables 6 and 7). Moreover, the experimental manipulation is not predicated on the tangibility of the theoretical entity because it is the independent variable, not the theoretical entity that is being manipulated. The independent variable is manipulated as a means to allow the hypothetical structure to show its properties or characteristics. In short, the observable consequences of unobservable entities can be used in a theoretically informed way to ascertain the tenability of the theory. A similar position is adopted in contemporary practice of psychological measurements (see the *Construction and Use of Psychological Tests and Measurements*).

6.2. Objectivity Despite Unreliable Perception

Phenomena of visual illusions and constancy are indeed threats to objectivity if experimenters rely on subjects' absolute performance. However, as may be recalled from Table 1, experimental conclusions are based on the differential performance between two (or more) conditions, not the absolute performance level in just one condition.

At the same time, despite its unreliability, our sensory system is quite stable. That is, the unreliability suggested by visual illusions or constancies does not change from moment to moment. Hence, if the comparison baseline is valid, the distortion due to the unreliability of our sensory systems would be the same in both the experimental and control conditions. In a sense, the unreliability is an extraneous variable that happens to be identical in the two conditions. Consequently, the unreliability can be excluded as an explanation of the data.

As visual illusions may be affected by the perceptual context, may objectivity not be compromised if manipulating the independent variable brings about a change in the context (i.e. context becomes a confounding variable)? If this possibility is known before data collection, and if the theory is specific enough, the experimental prescription would be in terms of the interaction between the independent variable and the context. Alternatively, the theory may be corroborated with a different independent variable that does not interact with the context. In short, the judicious application of deductive and inductive logic makes it possible to achieve objectivity despite the unreliability of the perceptual system under some circumstances.

6.3. Objectivity Despite Theory-Dependent Observations

Objectivity has been challenged by relativism. At the global level, social constructionists argue that the world as we understand it is our conceptual construction. The world would be different had we adopted an entirely different perspective. At the same time, the experimental hypothesis is what it is by virtue of what the research hypothesis says (see Table 6). If relativism is true at the global level, it may also be true at the level of individual experiments. That is, observations are inherently theory-dependent. Would experimenters not perceive only what is stipulated in the theory? How is objectivity possible under such circumstances?

Objective theoretical discussion is characterized by having an independent set of criteria to assess the acceptability of the evidence. That these criteria are available may be illustrated with the experimental corroboration of the iconic store. It is true that the experimental hypothesis of partial-report superiority is predicted on the substantive hypothesis of a transient, large-capacity visual buffer. It is also true that the dependent variable, the number of items available, is dependent on certain methodological assumptions. However, the assumptions underlying the identity of the dependent variable are not the substantive theory of the visual buffer. That is, the to-be-measured "number of items available" is a behavioral index that is independent of the theory of iconic store or any theory about the initial stage of visual information processing.

It may be argued that mental phenomena as we know them would be different had we adopted an entirely different worldview. For example, the "perceive more than can be recalled" phenomenon discussed in Appendix 1 might be called "XYZ" had we adopted a different worldview. A new theoretical entity (e.g. Agent W) might be required to deal with XYZ. That is, relativists may be claiming that even the to-be-explained phenomena themselves are theory dependent. However, relativists are unjustifiably conflating objectivity with absolute truth. If the purpose is to test the tenability of Agent W within the new wordview, and given the rationale of experimentation described in Tables 6 and 7, objectivity is not affected at all. That is, the logical foundation of the theory corroboration is indifferent to the underlying worldview in which it is embedded.

6.4. The Social Psychology of the Psychological Experiment

What may really threaten objectivity is the putative social psychology of the psychological experiment. The essence of this threat is the fact that conducting research is a special kind of social interaction. It has been suggested that three social factors work against objectivity: the experimenter expectancy effects (EEE), the subject effects, and the demand characteristics (for the importance of this issue in non-experimental research, see *Interviewing and Observation*).

6.4.1. The Experimenter Expectancy Effect

It is the EEE claim that experimenters, for various reasons, behave (intentionally or involuntarily) in a way that would produce data consistent with their vested interests. However, the bulk of empirical evidence cited in its support consists of showing how non-experimental data (i.e. data collected without any experimental control) might depend on who conducts the interview or collects the measurement data.
The apparent validity of the EEE claim comes from two experiments whose design is shown in panel 1 of Table 9. In the better known of the two studies, there were (a) the investigator (the one who set out to demonstrate EEE), (b) the data collectors (A, B, C, M, P, and Q, who administered the photo-rating task), and (c) the photo-raters (individuals whose responses became data). The investigator instructed the data collectors to replicate a photo-rating study.

Table 9. The distinction between the formal structure of the experiment (panel 1) and that of the meta-experiment (panel 2)

Panel 1. The formal structure of the experiment

Investigators					
+5			*–5*		
A	**B**	**C**	**M**	**P**	**Q**
S_1	S_1	S_1	S_1	S_1	S_1
…	…	…	…	…	…
S_n	S_n	S_n	S_n	S_n	S_n
$\bar{X}_A$	$\bar{X}_B$	$\bar{X}_C$	$\bar{X}_M$	$\bar{X}_P$	$\bar{X}_Q$
	$\bar{X}_{+5}$			$\bar{X}_{-5}$	

A, B, C, M, P, and Q are data collectors, not experimenters.

Panel 2. The formal structure of the meta-experiment

Investigator									
+5					*–5*				
D			**F**		**H**			**K**	
C	E		C	E	C	E		C	E
S_{C1}	S_{E1}		S_{C1}	S_{E1}	S_{C1}	S_{E1}		S_{C1}	S_{E1}
…	…		…	…	…	…		…	…
S_{Cn}	S_{En}		S_{Cn}	S_{En}	S_{Cn}	S_{En}		S_{Cn}	S_{En}
$\bar{X}_{(E-C)D}$			$\bar{X}_{(E-C)F}$		$\bar{X}_{(E-C)H}$			$\bar{X}_{(E-C)K}$	
$\bar{X}_{(E-C)+5}$					$\bar{X}_{(E-C)-5}$				

C = Control condition
E = Experimental condition
A, B, M, and Q are experimenters

The task for the data collectors was to present a set of photographs to a group of photo-raters. On the basis of the photographs, the photo-raters were to judge whether the individual in the photograph had recently been successful or unsuccessful by assigning a rating between +10 and –10. The expectation manipulation consisted of leading the two groups of data collectors to expect a mean rating of +5 and –5, respectively. EEE was said to have been substantiated because the +5 group of experimenters obtained a higher mean rating than the –5 group.

It is true that the study depicted in panel 1 of Table 9 is an experiment to the investigator. However, note that A, B, C, M, P, and Q collected data in one condition only. Hence, they were not experimenters for the simple reason that they did not have an experiment to conduct. They just collected measurements. Consequently, data from the oft-cited photo-rating study (or any study that has the same structure) cannot be used to substantiate EEE.

In order to substantiate EEE, individuals given different experimental expectations must have an experiment to conduct, as is the case with D, F, H, and K in panel 2 of Table 9. Consequently, to the investigator, the study should be a meta-experiment (i.e. an experiment about the experiment). The crucial comparison is the difference between the difference between the mean of the +5 group and that of the –5 group. The substantiation of EEE requires a significant difference between the difference between the experimental and control conditions. However, such was not found when the meta-experiment was conducted to test the EEE claim with the design shown in panel 2 of Table 9.

6.4.2. The Subject Effect

Critics of experimental psychology observe correctly that most experiments are conducted in colleges and universities. They make the point that not only is there no random selection of subjects, the representative nature of the samples is also suspect because students are different from the population at large. However, it is too simplistic and misleading to say that college and university students are not representative of the adult population at large simply because they are college students. For example, they too have four limbs, two eyes, and one mouth, like everybody. Depending on the frame of reference, students are different from, or the same as, the population at large. In other words, it is not meaningful to talk about representativeness in absolute terms.

Experimental psychologists do use non-student subjects when required. For example, they use hyperactive boys when they study hyperactivity. Moreover, even when they use student subjects, experimental psychologists do select their subjects in a theoretically informed way. For instance, they may use only right-handed subjects when they study hemispheric specialization or laterality effects.

In sum, although they do not always select their subjects randomly, experimenters do not thereby jeopardize the representativeness of their samples. Moreover, the crucial randomization procedure is the random assignment of subjects to test conditions. Experimenters routinely do this when the repeated-measures design is not used.

6.4.3. Demand Characteristics

To critics of experimental psychology, the demand characteristics of the experiment are the various personal and social pressures exerted on experimental subjects to produce data that are consistent with the experimenters' expectations and vested interests. What is the evidence for demand characteristics?

Evidence in support of demand characteristics consists of observations that, when the request was given as "experimental instructions," (a) thirsty participants continued eating salty crackers, (b) participants persevered in a boring task, (c) a research assistant was prepared to throw a beaker of concentrated sulfuric acid at another person, and (d) volunteers were prepared to comply with the instruction to grab a poisonous snake with a bare hand. The contention is that, if the participants are willing to do absurd, dangerous, or harmful things when instructed by the "experimenter," what else would they not do at the behest of the experimenter.

These examples suffice to show the absurdity of the "demand characteristics" critique of experimental psychology. It is misleading and incorrect to use non-experimental anecdotes to question the validity of the experimental approach. Suppose that experimental subjects were to ingratiate themselves to experimenters, they might succeed by modulating the absolute level of their performance. However, as has been shown in *Section 6.4.1. The Experimenter Expectancy Effect*, experimental evidence is inevitably in the form of the differential performance between the experimental and control conditions. Is it possible for subjects to produce the differential performance required?

6.5. Ecological Validity Revisited

As shown in Appendix 1, the partial-report task is a very unusual task. We seldom, if ever, have to respond to stimuli in the manner required. The example is typical of experimental tasks in that it is artificial and bears no resemblance to everyday phenomena. Hence, experimental data have been criticized on the grounds that they do not have ecological validity.

It is shown in row 14 of Table 5 that the importance of ecological validity differs for utilitarian and theory-corroboration experiments (for the "ecological validity" issue in non-experimental research, see *Interviewing and Observation*). That ecological validity is irrelevant, if not harmful, to theory corroboration may be seen with reference to Table 10. The to-be-explained phenomenon is that the ring road encircling the city is wet one morning. An obvious explanation is that it rained overnight. However, city employees might have washed the entire city. Hence, two alternative explanations are recognized in the table.

Table 10. The implications of two theories used to choose between the two theories: A case against ecological validity

1	To-be-explained phenomenon	The ring road around the city is wet		
2	Explanatory theory	It rained		City employees washed the city
3	Common implications of the theory	Wet ring road Wet side streets Wet footpaths Wet lower part of lamp poles		
4	Unique implication	Wet leaves on the top of tall trees		Dry leaves on the top of tall trees

If *It rained* or *City employees washed the city* is an acceptable explanatory theory, the least it has to do is to be able to explain additional phenomena other than the wet ring road. That is, any acceptable explanation must have multiple testable implications, for example, those shown in panels 3 and 4 of Table 10.

The four phenomena shown in panel 3 can be explained with both theories. Shown in panel 4 are two phenomena that can be explained only by either one or the other theory. *It rained* implies *Wet leaves on the top of tall trees* in the sense that *It rained* cannot be true as an explanation of the wet ring road if the leaves on the top of tall trees are not wet. Similarly, the phenomenon *Dry leaves on the top of tall trees* is an implication of the theory *City employees washed the city*.

In terms of the ecological validity argument, the phenomenon of wet side streets is the closest to the wet ring road. However, wet side streets are not useful as the criterion for choosing between the two theories because it is implied in both theories. At the same time, leaves are different from roads. Hence, subscribing to the ecological validity argument, one might not even think of checking the leaves on the top of tall trees when testing the two theories. Such data do not have ecological validity. Yet, one has to examine the ecologically invalid evidence in order to choose between the two theories because wet leaves is unique to one of the contending theories. In other words, a theory proposed to account for a common everyday experience is to be substantiated with data that are not similar, let alone identical, to the phenomenon in question.

Glossary

Complete factorial design: A factorial design all of whose treatment combinations provide data.
Conditional syllogism: A syllogism whose major premise is a conditional proposition.
Confounding variable: An extraneous variable found to have varied systematically with the independent variable of the experiment.
Constancy of condition: The experimental control achieved by (a) using the predetermined levels of the independent variable consistently and (b) holding the control variables constant.
Control: A provision in the experiment for excluding an alternative interpretation of the data.
Control variable: A variable that is held constant in the experiment. For example, *stimulus duration* is a control variable when the same duration is used at all levels of the independent variable.
Criterion of falsification: A theoretical prescription that has to be met in order for a theory to be tenable. A theory is declared false if research data do not match the theoretical prescription.
Deductive logic: A set of rules, if followed correctly, that would render an inference valid.
Dependent variable: The variable that is being measured in the experiment.
Experiment: An empirical study in which there are two or more conditions that differ in only one aspect. This is achieved by instituting all recognized controls.
Experimental design: The formal arrangement of the independent, control, and dependent variables of the experiment with reference to an inductive rule.
Extraneous variable: A variable that is not an independent or a dependent or a control variable of the experiment.
Factorial design: A multifactor design in which every level of an independent variable is combined with every other level of the other independent variables.
Hypothetical-deductive method: The rationale of scientific investigation introduced by Karl Popper. "Hypothetical" refers to scientists' speculation about why a phenomenon occurs or why the phenomenon is what it is. "Deductive" refers to the stipulation that the criteria of falsification of the speculation are its theoretical implications in various specific situations. That is, scientists make explicit what should follow from the speculation when a well-defined experimental manipulation is carried out in a specific set of conditions.
Incomplete design: A factorial design in which one or more treatment combinations is not used as the test condition.
Independent variable: A variable manipulated by the experimenter. This is done by setting up test conditions in terms of the levels of the independent variable.
Inductive logic: A set of rules for ascertaining the functional relationship between two variables by excluding alternative interpretations.
Method of difference: One of J.S. Mill's canons of induction. It is the inductive foundation of the experimental design in which there is an experimental and a control condition, as well as a set of theoretically recognized control variables.
Modus ponens: The deductive rule that says that affirming the truth of the antecedent of the conditional proposition warrants affirming the truth of its consequent.
Modus tollens: The deductive rule that says that denying the truth of the consequent of the conditional proposition warrants denying the truth of its antecedent.

Random assignment: Experimental subjects are assigned to the test conditions with a random process.
Treatment combination: A test condition that is defined by the combination of specific levels of two or more independent variables.
Validity: The logician's terms for "correctness."
Variable: Anything that can be represented in more than one way (or has more than one level).

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.]

Appendix 1. Whole-report and partial-report tasks

One reason that theory-corroboration experiments are not as readily comprehensible as utilitarian experiments is the conceptual distance between the instigating question and what is done in the theory-corroboration experiment. Hence, a strict discussion of *Section 5. The Theory-Corroboration Experiment*, particularly Tables 7 and 8, requires a description of (a) the two experimental tasks involved, namely, the whole-report task and the partial-report task, and (b) the theoretical reason behind introducing the partial-report task.

The whole-report task is more commonly known as the attention-span task. Subjects are shown a set of items very briefly (e.g. for 100 ms) and asked to recall as many items from the set as possible. The usual procedure is that (a) the series of trials begins with one item (i.e. set size is one), (b) the number of items in the set increases by one if subjects recall all the items correctly, and (c) the series stops as soon as subjects fail to recall all items in the set correctly. As the task is to recall all items, it may be characterized as the "whole report." The size of the last set that is recalled correctly is the subjects' attention span. Set sizes smaller (or larger) than the attention-span size is called the sub-span (or supra-span) size.

Typically, the subjects' attention span bears a 1:1 relationship with the set size when the set size is increased from one to up to four or five. However, the attention span reaches the asymptote beyond that limit in the sense that the number of items recalled remains constant at further increases of the set size. This asymptotic attention span was originally interpreted to mean that subjects could see, at most, four or five items at a glance. However, subjects invariably report that they have seen more than is indicated by their attention span. They also feel that, *had they been given more time*, they would have been able to recall more than the attention span.

Important to the discussion are two temporal factors. Recall the duration of the stimulus (typically 100 ms). However, subjects take typically one second to recall four or five items. Hence, subjects are performing in the absence of the stimulus for 900 ms of the time. This suggests a memory mechanism. The italicized *had they been given more time* refers to the duration of the content of the memory involved, not of the stimulus. That is, the limit of the attention span seems to be set by the duration of a memory component, not the amount of information that has been registered.

What is needed is a task that reflects the amount of information available at a glance without the *had they been given more time* limitation of the whole-report task. The partial-report task, which satisfies this requirement. Subjects are shown three rows of three items for 50 ms. A tone is then presented at the immediate offset, or at some delay after the offset, of the stimulus. The interval between the offset of the stimulus and the onset of the probe-tone is the "inter-stimulus interval" or ISI. Subjects are to recall **only** the top, middle, or bottom row of three items only. The tone is the probe-tone, and the selection criterion is spatial information (by row).

Note that the demand of recalling three items can be met because it is a sub-span set size. At the same time, subjects must have available all items in order to do the task because they do not know which of the three rows to recall until after it has been withdrawn. Hence, the estimated number of items available at the immediate offset, or any delay after the offset, of the stimulus is obtained with Equation [E1]:

The number of items available = The average number of items recalled per row × the number of rows (1)

The control condition is the whole-report task. Another control condition is having categorical membership as the selection criterion. Specifically, subjects are required to recall the letters if the probe-tone is high, symbols (e.g. %, #, etc.) if the probe-tone is medium, or digits if the probe-tone is low. Of interest is the difference between the mean "number of items available" in the partial-report condition and the mean number of items recalled in the whole-report condition. The typical finding is that the former is larger than the latter if subjects recalled by row (i.e. spatial information); this finding is called the "partial-report superiority." The partial-report superiority declines systematically as ISI increases; it disappears altogether by 0.5 seconds. Moreover, partial-report superiority is observed only with the physical criteria (color, size, and brightness), but not with category membership. This finding is the basis for accepting the view that information in the sensory store is "sensory" or pre-categorical.

It must be emphasized that the partial-report task is very unusual and difficult. To perform the partial-report task properly, subjects have to refrain from processing the stimulus until the probe-tone is presented. This means effectively not doing anything in the stimulus's presence. As this passive mode is against the subjects' habit, it is necessary to give them extensive training before collecting data. Specifically, it takes a large number of trials (100 or more) to render the data usable. At the same time, a large number of trials is also required. However, naive subjects would treat the task as the whole-report task: to start processing the items as soon as they are presented. For this reason, there is a disagreement about the visual sensory store because some investigators have failed to give their subject sufficient training or a sufficient number of trials.