PRA Inc.
Research & Consulting
www.pra.ca
admin@pra.ca

LABORATORY AND FIELD EXPERIMENTS IN ECONOMICS: SURVEY AND ASSESSMENT OF THEIR POTENTIAL FOR POLICY ANALYSIS

February 2008

Prepared for:

Human Resources and Social Development Canada

TABLE OF CONTENTS

1.0 Introduction 1
- 1.1 Purpose of the paper 2
- 1.2 Synopsis of the paper's findings 3
- 1.3 Guide to the report 3

2.0 Economic laboratory experiments: Context 3
- 2.1 Origins of experimental economics 4
- 2.2 Experiments as policy development tools 5
- 2.3 Clinical trials and internal/external validity 7
- 2.4 Economic laboratory and field experiments within the policy testing "landscape" 9
- 2.5 The example below shows how statistical and experimental methods may both address policy questions. 15

3.0 Issues in experimental economics 16
- 3.1 Theory validation 16
- 3.2 Experimental control 17
- 3.3 Internal and external validity 18
- 3.4 Salience and coherence 19
- 3.5 Financial incentives 19
- 3.6 Summary on experiments in economics 19

4.0 Case studies in laboratory and field experiments 20
- 4.1 Individual motivation 20
- 4.2 Eliciting time and risk preferences 21
- 4.3 Short- and long-term time preferences 22
- 4.4 Linking behaviour to program outcomes 23
- 4.5 Social capital experimentation 24
- 4.6 Experimental program evaluation 25
- 4.7 Targeted educational support policies 26
- 4.8 Using choice experiments in estimating the welfare wall 27
- 4.9 Using stated choice experiments to design an income support package 28

5.0 Summary 29
- 5.1 Final synthesis: How laboratory, field, and social experiments differ 29
- 5.2 Conclusion 31

Bibliography 32

1.0 Introduction

The word "experiment" conjures up images of laboratories, complex machinery, and people in white coats working on research projects far removed from our everyday experience. However, we all conduct informal experiments. As consumers, we routinely use trial and error to identify preferred brands of food, restaurants, and other daily items of consumption. Certainly, we do not subject foods to the "Pepsi" challenge, but trial and error, as an informal experiment, remains a major technique for deciding what to purchase.

Trial and error is time-consuming, and few of us have the patience for anything more complex than informal shopping around. We may improve our decision-making by consulting friends or reading product reviews. This additional information acts as a control on unknown factors that affect our ability to make a good decision. The more expensive the consumer product, the more likely we will take the time to consult additional sources of information or try alternatives. However, as consumers, we rarely subject our decisions to a blind taste test like that used in the famous soft drink ad. We simply do not have the time to evaluate products using a formal experiment.

Our high school science classes where we combined chemicals to produce an array of reactions are closer to a formal experiment. It was probably in that context when we first began to understand the concepts of controlled manipulation to study cause and effect. Simply combining chemicals haphazardly will usually produce no result; sometimes it may produce a flamboyant outcome. The high school experiment was controlled in the sense that the addition of a chemical produced an outcome that we compared to a baseline or reference state. Such experiments were also typically confirmatory, where the goal was to confirm a theory or test a hypothesis learned in a lecture. Experiments can also be exploratory to reveal new knowledge, especially when unexpected results occur. This leads to the standard definition of an experiment as an "action or operation undertaken in order to discover something unknown, to test a hypothesis, or establish or illustrate a known truth" (OED, 1980). This definition encompasses the following elements of an experiment:

- Manipulation of selected and isolated components (treatments) in the physical and/or social environment;
- Observation of changes in that environment before/after or with/without the manipulation; and
- Control of any factor that could change the outcome except for the treatments.

The focus of an experiment is usually to determine or discover cause and effect. The term "isolated" implies that the only allowable change is in a selected component, the *treatment* a process also known as the control of confounding factors. The term "environment" refers to the totality of physical and social experience, although for any experiment, the researcher may study changes only for a small segment, such as certain chemical agents or for non-human subjects placed in a specific type of situation. Such specific and limited changes in the environment are termed the *outcomes of interest*.

Traditionally, scientific experimentation has been the preserve of the physical and medical disciplines. A physical setting such as a laboratory provides extensive control over external or *confounding variables or influences*. This allows the researcher to analyze the change in the

environment in the presence or absence of the manipulation of a component, often referred to as the *treatment*.

In recent years, social scientists have adopted experimental procedures that attempt to replicate the control used by experiments in the physical and medical sciences. In addition to testing theories about human behaviour individually and in groups, this new wave of experimentation presents important opportunities for policy research and as a line of evidence in evaluation studies by HRSDC.

1.1 Purpose of the paper

This review explores the potential use of laboratory and field experiments to support social and economic policies and programs such as:

- training and employment support programs and services;
- policies that use financial incentives to encourage investments in various forms of social and human capital; and
- income support and redistribution programs directed at target clients including youth, Aboriginals, low-income families, and persons with disabilities.[1]

Formally, the terms of reference for this study are to "provide an overview of economic laboratory experiments and how this research tool may be used by HRSDC to inform its policy and program design and development, including performance measurement, and its partnership arrangements (e.g., third party delivery arrangements)." Though this study does provide a broad survey of how laboratory experiments may inform HRSDC policy and programming, it also demonstrates that the method and application of laboratory experiments has not been widely used in the social policy field. As a result, this paper also examines and emphasizes an evolution of the laboratory experiment, namely, the *field experiment.*

The following example typifies the laboratory experiments with which readers may be familiar. It illustrates how researchers use these experiments to test and refine economic propositions.[2]

Example 1: A simple laboratory experiment

The researcher approaches two participants (not necessarily selected at random), and offers the following proposition:

Participant A will receive $10 and can choose whether or not to offer part of that amount to participant B. B can choose whether or not to accept the offer.

Classical economics, based on rationality and selfishness, predicts that A should only offer a minimal amount — say $1 — and that B will accept, since something is better than nothing. In reality, when the game is played in the laboratory, B will reject any unfair offer, and A, anticipating this, offers more than a token minimum, typically $4. This experiment represents much of the recent work that refines fundamental axioms of classical economics. As modern economic theory stresses, rational self-interest also includes altruism.

[1] Throughout this paper, we will use the term "social policy" to refer to these types of interventions.

[2] See Lehrer (2006) for an accessible account of recent experiments in neuroeconomics.

1.2 Synopsis of the paper's findings

This paper argues that laboratory and field experiments offer a promising method for analyzing traditional questions concerning optimal policy and program designs and attributes of service delivery, and, under certain conditions, they can provide meaningful "evidence" on impacts.

This paper illustrates that laboratory and field experiments can be represented as opposite poles on a spectrum. At one end, laboratory experiments typically use convenience samples (undergraduate students in a university setting) to play games on paper and, increasingly, computers to explore and test basic economic theories of time preference, bargaining, and game theoretic. The related concept of testing individual preferences and choices using these experimental techniques can be taken further into the field, in what this paper calls field experiments. These experiments increase the realism or external validity of the research by setting the experiment in a more natural setting with participants that represent target decision-makers. At the extreme end of the spectrum, field experiments may encompass techniques common to business and market research, such as discrete choice (choices between two or more discrete alternatives) and conjoint methods (testing combinations of a limited number of attributes to determine which are most preferred by subjects).

Purists may complain that these market research studies, typically completed in the context of survey or focus groups, are not true field experiments. However, this paper extends the concept of field experiments to these common market research settings and techniques simply because they use methodologies, such as focus groups and surveys, common to applied research and evaluation studies undertaken by HRSDC. The benefit is that public policy research can be informed by these research methods at a relatively low cost.

1.3 Guide to the report

The remainder of this paper consists of four sections. Section 2 provides context regarding the use of economic laboratory experiments and its extensions, and discusses the use of experimentation for basic and applied research policy and program development as well as evaluation. Section 3 outlines some basic principles of this kind of experimentation, while Section 4 highlights previous experimental works and their potential applicability to policy development and evaluation. Section 5 summarizes the main points regarding this research method's experiments and implications for policy and program design and delivery.

2.0 Economic laboratory experiments: Context

Economics traditionally uses classical economic precepts such as utility and profit maximization to develop predictions of human behaviour. Empirical methods, typically using statistical methods and quasi-experimental designs, attempt to test these propositions.

In Canada, social experiments have been used to test certain hypotheses. These social experiments conform to the common medical sciences design. Treatment and control groups, usually chosen through random selection, define the social environment used to track outcomes arising from the selected application of an intervention (policy or programs). The Manitoba Basic Annual Income Experiment (MINCOME) and the recent work of the Social Research Demonstration Corporation are two examples of this approach.

Social experiments can be of small or are large scale and may be relatively expensive (see Section 2.4.3 below). Laboratory and field experiments are emerging as an alternative and lower cost way to test hypotheses about human behaviour and reactions to policy and program change.

2.1 Origins of experimental economics

After a protracted period of development, marked by some suspicion and even hostility, the award of the 2002 Nobel Prize in economics to Vernon Smith and Daniel Kahneman marked the acceptance of experimental methods in economics. In making the award jointly also to Kahneman, a psychologist, the 2002 Nobel Prize marked a milestone in the convergence of economics and psychology. This process had begun decades before and had been punctuated by the observations of J.M. Keynes on the marginal propensity to consume and the paradox of thrift, the Friedman-Savage hypotheses about insurance and lottery purchases, the consumption theories of K.J. Lancaster regarding the decomposition of goods and services into intangible attributes, and game theory with its strategic bargaining theories. The award to Smith, an economist, recognized the value of small-scale laboratory experiments in exploring the validity of predictions about human behaviours.

The 2002 Nobel Prize marked a formal acceptance of experimentalism by economics, which traditionally has used observations from real world data to infer causal relationships. Previously, most economists collected data from objective sources, such as the stock market, or from subjective sources, such as sample surveys. In fact, many prominent economists from J.S. Mill to Milton Friedman to Paul Samuelson disavowed experimental methods and stated, "Economists cannot perform control experiments of chemists or biologists because they cannot easily control other factors."[3] However, over the last few decades, an increasing number of economists have accepted that data resulting as the outcome of experiments are as equally legitimate as data from more conventional observational sources.

In other settings, experiments have formed a basis for scientific inquiry for hundreds of years. The process of testing theories either confirms or refutes current doctrine, thereby solidifying confidence in a theory or leading to its demise. For example, the Michelson-Morley experiment that showed the speed of light as constant in all directions dealt a fatal blow to the ether theory of light and became an instrumental fact that led to the theory of relativity and quantum physics.

It is important to emphasize that science has progressed on the basis of both observational and experimental studies. The theoretical revolutions of Galileo and Darwin are examples of inferential leaps based on observations. Until recently, economics used the observational approach alone to develop and validate theories. Not only have experimental methods considerably broadened the techniques economists have to validate theories, but also the key idea that merits reiteration is that they can be used to *generate data to serve as the basis for new theory*.

Is experimental economics simply an extension of the natural science approach to experiments? Guala argues it is not, in that natural science tends to neglect the issue of external validity — the extension of the findings of an experiment outside the laboratory.[4] He also argues that

---

[3] As quoted by Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists.* Cambridge: Cambridge University Press.

[4] Guala, Francesco. (2005). *The methodology of experimental economics.* Cambridge: Cambridge University Press.

experimental methods in economics are far from finalized, and considerable development is occurring in approaches and practice. The applications of experiments to policy are also relatively sparse, at least within the economics profession. Yet, if one broadens the definition of experiment to include research that explores reactions of randomized groups to variations in policy and program design, an array of potential methods emerges.

2.2 Experiments as policy development tools

Early economic experimentalists in the 1930s and '40s worked in three distinct areas. Industrial organization theorists such as Edward Chamberlin were primarily interested in testing new-classical theories of market functioning.[5] A second strand of work focused on testing the implications of game theory.[6] The third studied individual decision-making using the theory of choice under uncertainty,[7] which culminated in increasingly complex experiments that characterize the Nash equilibrium, a position where no "player" has any incentive to make a move (cooperate, defect, etc.). As is evident from the standard references in the area, much of the current research simply elaborates on work in these three areas.[8] However, several recent publications have reported laboratory experiments in policy development. As examples:

- In an application to manage the current energy price escalation, Kruse, Elliott, Schulze, and Ben-David test policies for rationing supply in the face of external shocks.[9]
- An application to emissions trading appears in Ishikida, Ledyard, Olson, and Porter.[10]
- A policy application to agriculture and environment appears in Cummings et al., but this too is a policy application of an auction process.[11]
- Falk and Fehr make the case that experiments can complement traditional methods of economic analysis to support labour market policies.[12]

These selected policy applications reflect the roots of experimental economics in auctions, market operations, and individual valuations under uncertainty where laboratory experiments in economics remain focused. A scan of *Experimental Economics*, the main journal in the field, reveals few applications to policy design that easily falls within the scope of interest of HRSDC.

---

5 Chamberlin, E. (1948). An experimental imperfect market. *Journal of Political Economy, 56*, 95–108.

6 The standard example of this is the "prisoner's dilemma." Tucker, Al W. (1950). *A two-person dilemma*. Mimeo. Stanford University. Published under the heading "On jargon: The prisoner's dilemma." *UMAP Journal,1* (1980), 101.

7 von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behaviour.* Princeton: Princeton University Press.

8 Kagel, J.H., & Roth, A. (1995). *The handbook of experimental economics.* Princeton: Princeton University Press; Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists.* Cambridge: Cambridge University Press; Davis, D., & Holt, C. (1993). *Experimental economics*. Princeton: Princeton University Press.

9 Kruse, Jamie Brown, Elliott, S., Schulze, W., & Ben-David, S. (2001). Rationing supply capacity shocks: An experimental comparison of practical curtailment mechanisms. In R. Mark Isaac (Ed.). *Research in experimental economics, Vol. 8* (pp. 153–183). Greenwich, CT: JAI Press Inc.

10 These two papers appear in Isaac, ibid.

11 Cummings, Ronald G., Holt, Charles A., & Laury, Susan K. Using laboratory experiments for policy making: An example from the Georgia irrigation reduction auction. Department of Economics. Andrew Young School of Policy Studies, Georgia State University. http://www2.gsu.edu/~wwwenv/waterPDF/W2002003.pdf

12 Falk, Armin, & Fehr, Ernest. (2003). Why labour market experiments? *Labour Economics, 10*, 399–406.

Even Guala,[13] arguably the most useful review of the foundation of experimental economics, does not cite policy applications of laboratory experiments as conventionally defined.

However, this situation is now changing. The game theory experiments of the 1960s and '70s are evolving into public goods experiments that attempt to predict how individuals will react to policies around crime, pollution, and global warming (public "bads") as well as national security, clean air, reserving land for wilderness, and public health.[14] As experiments move from the laboratory to the field, policy applications and opportunities are increasing. Before delving into this more deeply, it is worth standing back and placing laboratory and field experiments in a more general policy testing landscape. Since this paper deals with experiments in economics, it is useful to start with a review of the essential features of clinical trials to provide a backdrop for laboratory and field experiments in economics.

As a preliminary observation, it is worth noting that the creation of a treatment and control group using random assignment of subjects is less common in the laboratory and field experiments used in economic policy analysis. Randomization is nothing more than a procedure to control for external confounding factors. Laboratory and field experiments use other methods to control confounding effects in addition to randomization of subjects.

Random assignment and experimental control

Pharmaceutical research and the social experiments used in economic policy use random assignment to create treatment and control groups. It seems intuitive that this controls for external confounding influences, but to understand this it is necessary to explain the concept of statistical equivalence.

Imagine one were interested in creating a treatment and control group from the general population. Randomly selecting two individuals, assigning one to the control group and the other to the treatment group, would certainly create non-equivalent groups. The first member of the treatment group might be a woman, while the first member of the control group might be a man. Income, age, education, etc. would also be different. However, as one added subjects to each group (using random assignment of course), in terms of the proportion of men and women and the average age, education, income, etc. the two groups would be increasingly similar. Two key attributes of randomization confirm this as a technique to create statistical equivalence:

- As the sample size increases, the statistics one uses to compare the treatment and control group will become equal. The term "statistic" includes averages, variances, ranges, or any other measure of a sample.
- Not only does randomization cause the statistics of observed attributes to become equal, it also causes the statistics of unobserved (unmeasured and unknown).

Randomization creates statistically equivalent treatment and control groups on observed and unobservable attributes. Therefore, the introduction of a change to the control group will produce an outcome that is uniquely attributable to that intervention.

13 Guala, op. cit.
14 Ibid.

2.3 Clinical trials and internal/external validity

Clinical trials are the "gold" standard in any experimental work. These trials are specifically designed to test the relation between a cause, the drug, and the effect — a change in health status. Randomized, double-blind clinical trials are the prescribed method used to confirm a causal relation between the drug and the outcome. This section presents an example of a clinical trial for a new pharmaceutical to illustrate the essential principles of an experimental study.

Donezpil is a member of a recent class of pharmaceuticals designed to slow the decline of patients suffering from Alzheimer's disease. Since this condition can only be confirmed by autopsy, patients suspected of suffering from this disease usually participate in questionnaire-based tests designed to assess cognitive ability. Figure 1 shows a typical result of clinical trials used in the approval of these treatments (see Rogers et al. (1998) for a typical study). This example shows a control group and two treatment groups with participants randomly assigned to each group. The study is double-blind in the sense that participants (and their caregivers) as well as research and test administrators have no knowledge of who is receiving a pill with the active ingredient or placebo (a pill with no active ingredient). These trials confirm the efficacy of Donezpil, which has now become a standard therapy in the treatment of Alzheimer's. Note the distance in responses over time and how the experiment can clearly distinguish between them.

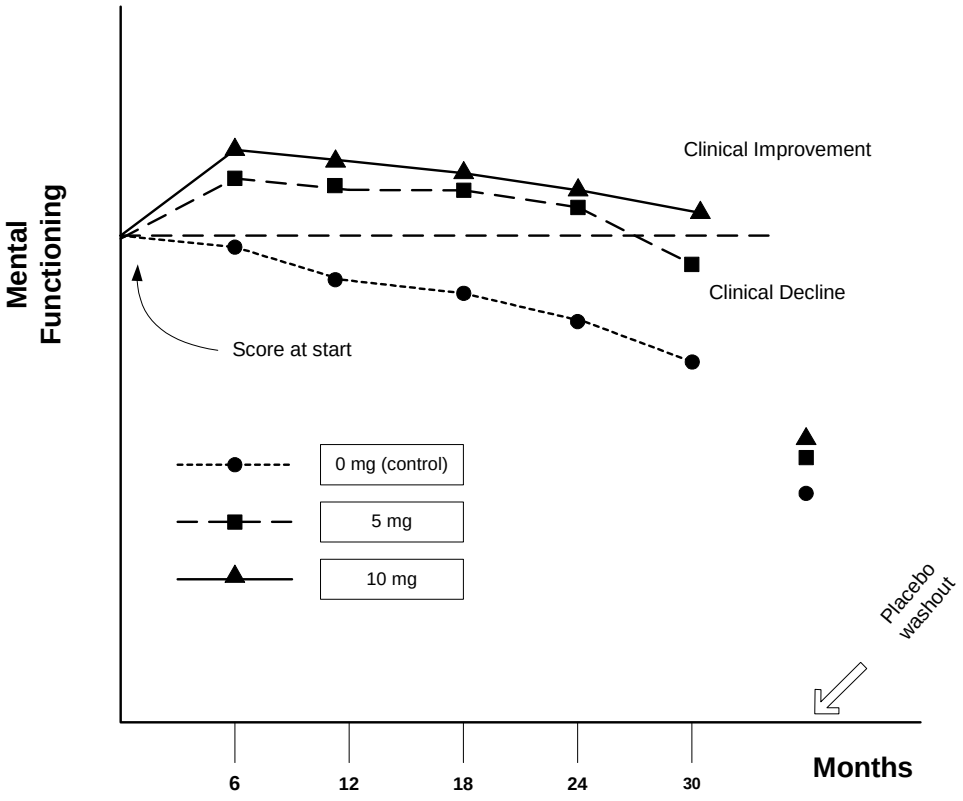

Figure 1: Typical clinical trial for Donezpil

Internal validity refers to the integrity of the procedures used to randomize treatment and control groups and the methods used to administer the treatment and record the outcomes. Internal validity refers to the logical links between the objective of the policy or program, the design of the treatment, and the expected effect. These should be internally consistent.

Internal validity also refers to the logistics of an experiment. For example, double-blinding means that those administering and those receiving the drug do not know who is getting the pill with the active ingredient or the placebo. This controls for sympathetic caregivers giving the active ingredient to those who are "deserving" or who have a better chance of benefiting. Ignorance on the part of the patient controls for the effects that optimism or pessimism might have on the outcome.

A well-controlled trial pertains only to those in the experiment. If the size of the treatment and control groups is sufficiently large and drawn from a diverse population, then the results of the trial may be generalized to a wider population. The study then has increased its *external validity*. For example, clinical trials in the laboratory using non-human subjects to first ascertain the potential for harm to humans are a prelude to field studies with humans subjects.

Another approach to increasing external validity is replication of the experiment to other settings and populations. Replication also validates the experimental results and serves as a check on internal validity and faking.[15]

A final feature of the above clinical trial is *treatment or dosage manipulation*. In this example, two treatment levels demonstrate that the higher dose has more effect, but not by much. Dosage variation confirms the action of the drug; however, it also identifies whether increasing the active ingredient has a benefit. This is important for a commercial reason, and also to determine whether unexpected and unwelcome side effects are associated with higher doses. The final treatment manipulation is the *placebo washout*, which places all participants in treatment and control group on the placebo. The convergence of all participants to the same low level of cognition confirms the action of the drug.

Clinical trials demonstrate the key features of a laboratory experiment — randomization as well as treatment/control administration protocols to ensure the integrity of the experiment *(internal validity)*, methods to support generalization to a wide population *(external validity)*, and dosage manipulation as an additional method to confirm the drug.

Economic and social policy and programs development generally operate within a causal analysis framework — the intervention should lead to a specific outcome. In Canada, policy-makers have relied most heavily on a range of theoretical processes *(*simulation and *thought experiments)* and statistical procedures applied to administrative and survey data principally that fall under the terms *non-experimental* and *quasi-experimental analysis*. There has been some use of social experiments and very little use of economic laboratory experiments to test for preferences or to learn more about how policy and program attributes influence individual choices. The next section reviews the policy research and testing landscape in economics and shows how laboratory and field experiments could fit within the terrain at HRSDC.

---

[15] Giles (2006) discusses the concept of replication in natural science experiments and observes that some aspects of sciences such as astronomy cannot conduct experiments. Conclusions on cause and effect must be inferred from observing a series of single events, such as past explosions of supernovae. This bears a close resemblance to the inferential problem faced by economists.

2.4 Economic laboratory and field experiments within the policy testing "landscape"

Social policy analysis and development rests on a theoretical and empirical understanding of cause and effect relationships in society. Figure 2 presents a schematic of the ways economists test policy. Each of these techniques is in current use. As mentioned in the previous section, thought experiments and non-experimental methods including quasi-experimental studies remain dominant.

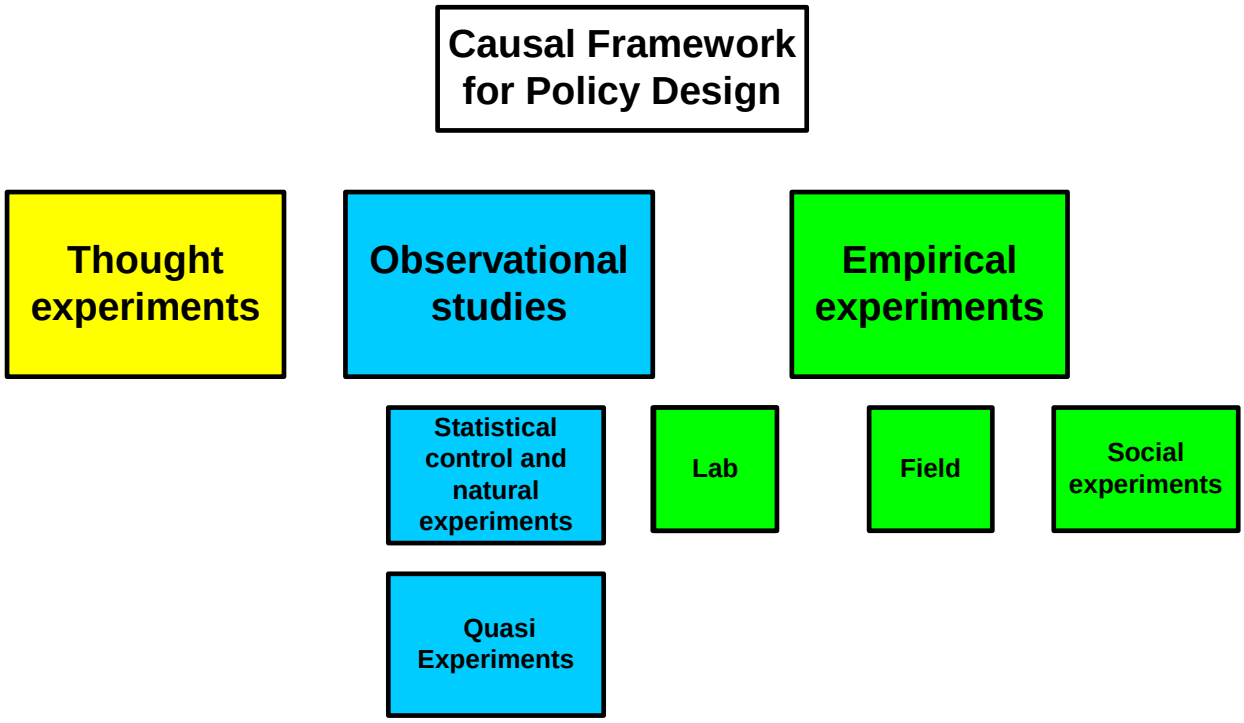

Figure 2: The landscape of empirical economic policy and program design/evaluation

2.4.1 Thought experiments

A thought experiment is a "what if" exercise that can be performed mentally, mathematically, or simulated using a spreadsheet or a computerized model to facilitate calculations and develop and appreciation of potential outcomes.[16] This process starts from axioms of behaviour and also possibly estimates of empirical relationships. This style of abstract modelling is widespread in economics, which typically starts with "let us assume…." Theorists manipulate equations and model potential causal relationships. (Einstein is possibly the most famous scientist to use this approach.) In today's environment, many models need to be and can be tested/solved numerically.

Much of social policy analysis proceeds on the basis of thought experiments, the simplest of which ranges from the shifting demand and supply diagrams common to first-year economics courses, to the more complex being dynamic modelling using advanced mathematics and

[16] See Gendler, T.S. (2000). *Thought experiment: On the powers and limits of imaginary cases.* New York: Garland. See http://en.wikipedia.org/wiki/Thought_experiment (although Wikipedia changes almost weekly).

computer simulation. These models, though elegant, are often unpersuasive, largely because the behavioural assumptions upon which they rest are not always tested and accepted. One of the important contributions of laboratory and field experiments is the evaluation of these behavioural assumptions through the testing of basic human behaviour when faced with complex decisions. The first example presented above illustrates how to test two basic assumptions of neoclassical economics — rationality and selfishness. It is probably not the case that individuals are simplistically selfish, but rather most operate within a more complex process of strategic action and reaction. Sometimes, people may offer more out of the goodness of their hearts, and at other times, the decision may be strategic in an attempt to ensure reciprocity when the "shoe is on the other foot." However, context matters and incentives matter. Without further knowledge about these influences, the thought experiments become overly speculative.

2.4.2 Observational studies (quasi-experimental methods)

This is by far the most common approach to policy and program evaluation. The backbone of this methodology in economics comprises simple pre-post comparisons in the time domain or comparing outcomes between a program and comparison group simultaneously. Estimates or net impact emerge from a multivariate analysis, or quasi-experimental methods where program and comparison groups track outcomes for those exposed and those unexposed to policy options. The following figure presents a stylized model of the multivariate analysis method (Figure 3). It is nothing more than a map of a regression model. The coefficients $B_1 - B_k$ estimate the strength of the association between the independent variables X (the causes) and the dependent variable Y (the effect).

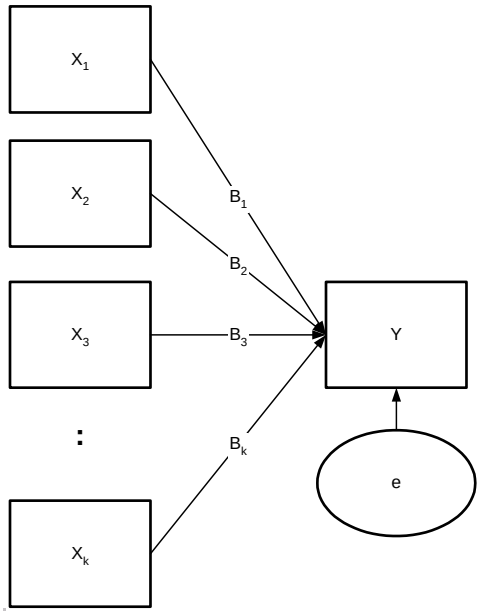

Figure 3: A map of a regression model

The simplest approach to modelling policy or program impacts is to develop one of the X variables as an indicator of the policy. This can be as simple as using a dummy variable that assumes the value of 1 for when/where/for whom the policy takes effect and 0 otherwise. Unfortunately, many view this approach to causal analysis as unpersuasive. Most important,

often the analyst must impose causal direction and must select the best candidates for the independent variables. Heckman situates the problem as follows:

> Some of the disagreement that arises in interpreting a given body of data is intrinsic to the field of economics because of the conditional nature of causal knowledge. The information in any body of data is usually too weak to eliminate competing causal explanations of the same phenomenon. There is no mechanical algorithm for producing a set of "assumption free" facts or causal estimates based on those facts.
>
> Heckman (2000)[17]

There are ongoing attempts to make the non-experimental and quasi-experimental approaches more rigorous because quite often these are the only methods available to social scientists to test the possible effects of interventions or for conducting program evaluations. Much of current econometrics focuses on the task of identifying causal relationships within the set of data at hand. Instrumental variables and simultaneous equations represent common approaches. Several methods in particular are noteworthy for this discussion.

- The ***instrumental variables*** method, developed in 1928 by Wright, remains the classic technique for extracting causal insight from observational data. The simple demand-supply model, with price and quantity as the only variables, provides insufficient information to identify behavioural relationships from market data. Wright's insight that using a third variable (the instrument) can trace out a behavioural relationship, remains a landmark in causal analysis in economics. It is interesting that the conceptual model remains very close to a pure thought experiment. Unfortunately, what dooms instrumental variables in most cases is the data; it is hard to find a good instrument (i.e., a variable that influences participation, but is not related in a significant way to responses / outcomes). In most cases, datasets available to the analyst does not include variables that determine participation, and which also are not correlated with outcomes. Some classic studies have been successful, but these tend to be relatively rare.

- ***Natural experiments*** exploit the opportunity presented by a policy change or unique event. These methods closely align with pre-post analysis and the multivariate models that use statistical control to estimate the effects of an intervention, where outcomes before and after or between the program and comparison regions may be modelled as a function of socio-economic variables and an indicator variable. Other methods, such as difference-in-differences (e.g., examining the pre-post conditions for both participant and non participant groups and comparing outcomes), may be used on the before and after time series for otherwise identical program and comparison groups. The classic example of a natural experiment is the work of Card and Krueger (1995) on the minimum wage. In that study, the authors compared teen employment, primarily in fast food restaurants, in two adjoining states (Pennsylvania and New Jersey) after one state increased the minimum wage. The results that showed no negative outcomes on employment reignited the debate on the employment effects of the minimum wage, after most economists believed that the matter had been settled. Recently Oreopoulos (2006) used a change in high school leaving age in England to estimate the returns to schooling. The key

[17] Heckman, J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *Quarterly Journal of Economics, 115*, 45–97.

challenge for a natural experiment is to show that no other differences exist between two states or over time that could account for any observed differences.

- ***Propensity matching*** has emerged as one way of supporting analysis to identify causal relationships generally and specifically as a method to estimate the net impact of programs and policies. Imagine that a program, such as labour market training, has participants and non-participants. Just comparing the outcomes for these two groups theoretically is wrong because observed and unobserved differences may exist. For example, non-participants may have less motivation in general or may have skills, which means it makes sense for them to prolong their search and not participate in their training. Propensity matching uses statistical methods in an attempt to create "statistical equivalence" between participants and non-participants based on observable characteristics. In effect, propensity matching tries to approximate a randomized experiment and hence its classification as a "quasi-experimental" technique. The box below explains propensity matching in a little more detail and can be omitted without loss in continuity.

Brief review of propensity matching

Matching processes attempt to create a treatment and control group from participants and non-participants respectively. Pairwise matching selects "matched pairs," one from the non-participants to form the control group and one from the participants to form the treatment group. It is usually easy to match on a single attribute such as gender, but to match on multiple attributes (age, ethnicity, region of residence, education, etc.) becomes progressively more difficult. One can loosen the match, matching on age ranges rather than the specific age, but this merely increases the statistical *in*equivalence of the treatment and control groups.

Propensity matching uses a statistical approach. The first step is to estimate the likelihood of participating in the program. Participants are assigned a score of 1 and non-participants a score of 0. Logit or some other probability model provides estimates of the probability of participation based on a range of independent variables. The result of this estimate process is that participants and non-participants alike will have an estimate probability of participation; non-participants will have lower scores than participants, but there will be overlap. Treatment and control groups are formed by matching participants and non-participants with similar probabilities.

The challenge faced by propensity score matching for creating comparison groups is to develop a good statistical model of participation (recreating the world without the program for participants). In reality, administrative data sets and specialized survey data tend not to offer sufficient information (i.e., there is missing information) to draw firm conclusions about the validity of the comparison group. Such studies, like the vast majority of social science research studies that use administrative and survey data, experience difficulties that are associated with data errors, recall errors, and refusal to respond to survey.

2.4.3 Empirical experiments

The key idea of experimentation is manipulation according to some rule and comparison of outcomes from various trials to some norm. In the clinical trial, that norm is the control group. Economics has used social experiments and more recently laboratory and field experiments to estimate the causal relationship underlying theory and to guide policy. The main experimental methods are described below:

Social experiments represent the economist's attempt to faithfully replicate the approach of the natural and medical scientist. These studies are often cited as the "gold standard" in economic policy evaluation. The *Job Training Partnership Act* in the United States mandated randomized trials to evaluate programs funded by Congress. In Canada, the work of the Social Research and Demonstration Corporation on the Self-Sufficiency Program, *learn*$ave and the Community Employment Innovation Project are examples of these large-scale studies. The MINCOME Manitoba study of the early 1970s was the first and, until the Self-Sufficiency Project, the only social experiment in Canada.

Random allocation of subjects into treatment and control groups, application of selected program alternatives to the treatment group, monitoring of behaviour change, and comparing the responses of the two groups provide a measure of net impact. Social experiments date from the mid-sixties and work continues; a recent reference is SRDC (2006). Social experiments can take many forms. They can be used to address questions of service delivery, as well as policy relevance and effects. The larger-scale projects dealing with complicated policy issues can take a considerable amount of time to design, operate, and produce results. Depending on the intervention, they can also be expensive. The major expense isn't the evaluation component. Often the payments that mimic the policies being tested comprise a very high percentage of the experimental cost.[18] These factors naturally lead policy-makers and researchers to consider alternatives for producing "evidence" on impacts in shorter and possibly less expensive time frames. This is one of the main factors influencing the emergence of, and recent interest in, the laboratory and field experiment.

In addition to the factors just mentioned, social experiments, particularly those of the larger scale variety, require the cooperation of many partners. In Canada, this typically involves two or more orders of government and various non-governmental organizations.

- ***Laboratory and field experiments*** are emerging from decades of relative obscurity. Harrison and List (2004) propose the following definitions:
  - ***Economic laboratory experiments*** use standard subject pools, typically university students, and abstract (somewhat artificial) settings, typically computer labs, and a set of rules for the experiment. The experiment often consists of a game or program "played" by the participants where rewards may not usually be tangible. The outcomes from laboratory experiments consist of choices made by players (participants) and the results (gains and losses). Experimentalists use this setting to explore various propositions from game theory, bargaining, and auctions. The obvious weakness of laboratory experiments is their external validity in that

[18] The expenses associated with a large-scale social experiment should be contrasted with the expense of a fully implemented program, which might ultimately be found to be less than or not effective.

university students are hardly representative. However, because the experimentalist can impose significant control on the context and the "game," internal validity is very high.

– ***Economic field experiments*** as extensions of the laboratory experiment represent an area of growing interest and relevance for policy development. The field experiment "game" is closer to reality than a laboratory experiment, with rewards and losses being significant. Rather than university students, for example, players may represent the target clients for a policy. A field experiment also may attempt to replicate the setting in which choices may be made. In this sense, field experiments in economic policy resemble trials in market research with consumers participating in choices in stores and restaurants. Realism may be increased in that playing the game well can mean important financial gains for participants. The use of monetary rewards is a feature of the economist's approach to experimentation, in contrast to that of the psychologist, where financial compensation is rare. Increasing the stakes is widely seen as increasing the relevance of the process to experimental subjects and aligning the results closer to real world outcomes. Within field experiments, Harrison and List identify two general approaches.

    - ***Artefactual field experiments*** apply the laboratory experiment to non-standard subjects. This may seem a small adjustment, but it increases external validity, at the expense of internal validity. Some researchers extend this one step, using heterogeneous subject pools and then multivariate analysis to model the impact on experimental outcomes of covariates such as age, gender, and education.

    - ***Framed field experiments*** move the artefactual experiment into a field setting, such as a store, online purchasing, or choices made in a bank. In an economic policy context, offering different policy options may be one way to frame an experiment. Exploring options in social housing, for example, may best be completed with residents in their units.

2.5 The example below shows how statistical and experimental methods may both address policy questions.

Example 2: Job search options[19]

Anecdotal evidence from evaluations of employment programs suggests that counsellor attention is an important factor in motivating and orienting unemployed persons seeking work. In a typical job and training support centre, for example, the effect of trained assistance on the speed with which the unemployed can generate job leads may be studied statistically or experimentally.

The statistical approach involves logging the time spent by applicants in generating five legitimate job leads, while observing the length of time applicants have access to trained counselling. Analyzing variation in the time spent generating job leads, through multivariate methods, provides an understanding of the factors that affect outcomes.

An experimental approach would introduce structure by selectively withdrawing and offering counselling assistance and observing the impact on the number of job leads. Clearly, this would require staff training and careful control to accurately vary the time and attention paid to clients.

The use of an experimental approach for this policy test may seem contrived, but several observations emerge.

- The experiment should be blind to the participants. For example, an experiment may apply differential counselling services to subjects by having counsellors "absent" and unable to deliver services to clients, ostensibly due to sickness, professional development, or other legitimate reasons. This would differentially apply services to experimental subjects, while not alerting them to the experimental process (simulating experimental conditions in pharmaceutical experiments) .

- It is important that the experiment do no harm. Job seekers might become discouraged by not having access to a counsellor and may not return. One approach may be to complete follow-up to recover participants who quit early, and the experiment should not last long. Another approach is to add staff to increase attention and time spent with clients so that it could be argued that no one received less service as a result of the experiment.

- Typically, evaluations of net impacts from training programs use changes in earnings and employment. Consider that government has invested and continues to invest in human resource centres to support job seekers. Measurement of the benefit of counselling typically uses qualitative methods (focus groups) and/or sample surveys, where respondents offer their impression of the assistance provided. Aside from the biases inherent in this representation, the outcome is often a vague recollection by participants months or even years after the fact. An experimental approach such as described is easy to implement, low cost, and would focus on a central assumption of these centres, namely, that in-person counselling has a net benefit on job search and employment recovery. It would also produce much faster results than typically available in evaluation studies.

[19] Examples without references illustrate a hypothetical situation.

- Finally, many experiments combine statistical analysis with experimental manipulation. This increases the control and external validity of the experiment. For example, including the attributes of job seekers as covariates along with variables that express the variation in time spent by counsellors can increase the insight into the factors that produce a successful job search, since they might not be the same for all.

Laboratory and field experiments have been proposed as an alternative to observational studies and social experiments. Before presenting examples of such experiments, it is useful to review the requirements for successful experimentation in economics.

3.0 Issues in experimental economics

It is possible to skip this section on first reading, but a deeper understanding of experimental economics requires an appreciation of the issues discussed in this section.

Five themes underpin experimental economics:

- theory validation
- experimental control
- internal and external validity
- salience and coherence
- financial incentives.

3.1 Theory validation

The essence of any policy/program evaluation is that the research should clarify causal relationships. The dominant paradigm is the falsification process proposed by Karl Popper asserting that knowledge (causality) emerges as the result of rejecting false propositions (null hypotheses). Consensus exists among scientists (natural and social) that experimental methods provide a stronger basis for rejecting false ideas than non-experimental methods. A laboratory supports well-defined environments (initial conditions) and the explicit isolation of interventions, essential for experimental control. Smith (1994), a pioneer in experimental economics, expresses this requirement as follows:

> All tests of a theory require various auxiliary hypotheses that are necessary in order to interpret the observations as a test of the theory. These auxiliary hypotheses go under various names: initial conditions, *ceteris paribus* clauses, background information, and so on. Consequently all tests of a theory are actually joint tests — that is, a test of the theory conditional on the auxiliary hypotheses.
>
> V.L. Smith (1994)

This issue lies at the heart of theory validation. All theories require a set of untested assumptions or "maintained hypotheses." It is impossible to control for everything. Strict followers of Karl Popper argue that there is no such thing as confirming a theory, and that truth emerges only by falsifying a successive and presumably endless set of propositions. We see this in classical statistics that present a null hypothesis. Strictly speaking, a significant statistical result only tells what the coefficient is unlikely to be, not what it is likely to be. The weakness in the falsification

perspective is that after rejecting a null hypothesis, in theory, an infinite number remain to be tested.

Of course, most scientists accept a more practical relaxed view that successive experiments allow one to induce the validity of a causal relationship through the replication of studies and experiments.[20]

Example 3: Experiment in altruism (public goods experiments)[21]

(This example can be omitted on first reading)

Economists have been fascinated with the social dilemma type game. The simplest is the well-known prisoner's paradox. Doing what is best for one's self often makes the situation worse for all. These games illustrate that selfish behaviour often places the group in a worse position. At the same time, varying the rules of the game can change outcomes substantially.

One variant of the game uses several players, each of whom is offered the following "deal." All players play at the same time and anonymously (a key feature of the game). Each player is given an initial stake of 20 tokens and told that he/she may keep all 20, or donate a portion to the group (to the public good). Every donation will be multiplied by two and shared equally by the group. For a four-person game, the payoff to person i is

$P_i = 20 - g_i + .5 \sum g_j$ (j from 1 to 4)

Researchers have noted that at the early phases of this game, players tend to make public investments, but over time, realize that keeping the money provides them with a higher return.

The process of replication with the same group of participants supports a sort of "triangulation" to an increased understanding of how individuals might behave in a series of sequential choice situations. The issue for experimental economics is that "truth" does not locate in the testing of one hypothesis, but through testing a series of linked hypotheses that may examined by making incremental changes to a "game." Alternatively, one can repeat the game with different groups. The key is that systematic replication produces increasing understanding.

3.2 Experimental control

The typical laboratory experiment in economics occurs in a laboratory setting with payments to participants to "play" the game. The "laboratory" may include a room where participants complete pen and paper exercises or "games" on a computer network. Experimentalists in the laboratory often pay little attention to randomizing subjects; rather, the focus is on controlling all confounding effects. Considerable effort goes into the clear definition of the rules of the game, ensuring that all players understand these rules and play by them. Most laboratory experiments in economics now use computer games, managed over a network. Some of the prominent researchers maintain websites where individuals may play these games.

---

[20] Note that repetition differs from replication. Repetition is the exact duplication of a study or experiment and serves only to check the integrity of the researcher. Replication is the reproduction of an experiment with small and known variations. If the data from the two experiments are generally in agreement, replication supports the generalization of the study and acquisition of new insight through a process of induction.

[21] From Guala (2005).

Laboratory experiments often dispense with randomization and replications with different populations or variations in the game to the *response database*. However, randomization along some dimension, to ensure experimental control, is often used with field experiments. In the example using counsellors for unemployed persons, randomizing on individual clients would not usually be possible. However, recent developments in place and time-slot randomization enable randomization in situations that typically would not support it.[22] Therefore, applying the different counselling approaches at randomly selected times and across randomly selected HRCCs introduces the randomization and would increase the internal validity of the experiment.

3.3 Internal and external validity

A field experiment often proceeds with a specially enrolled group that participates in a focus group room, or in a conventional survey (telephone, mail, or web), or in the setting where participants make decisions. The key idea is that the field replaces the laboratory and that using target clients increases the external validity. This presents one of the core dilemmas for economic experiments — a trade-off exists between internal and external validity.

Internal and external validity are desirable features of a laboratory and field experiment in economics. To recap:

- *Internal validity* refers to an experiment that eliminates confounding variables. An internally valid experiment supports the inference of causal relationships from the experimental results within the context of the experimental setting. Laboratory experiments typically have higher internal validity than field experiments, largely because the field setting loosens the control of the research on the participants and the nature of the experiment.

- *External validity* means that it is possible to generalize results from the laboratory to the wider community with some confidence. Field experiments, especially those that replicate the actual policy delivery setting, support much greater generalization from the test population. However, assuring high levels of external validity may weaken internal validity to the point where confounding factors influence and invalidate experimental results. To recall the clinical trial: Showing that a drug has an effect on subjects in a closely controlled setting has strong internal reliability; taking that experiment to a series of hospitals across the country increases the external validity, but the control over how the experiment is actually conducted in remote sites is correspondingly less.

Clearly, logistical control is important in any experiment, and those participating in the experiment must adhere to experimental protocols. Again, in the counselling example above, a threat to internal validity would occur if some counsellors override the experimental protocols and fail to apply different treatments as required. In contrast, a computer-based experiment where the rules of the game may be strictly enforced ensures high degrees of internal validity, yet results may not be readily applicable outside of the laboratory. This leads to the common observation that moving from laboratory to field reduces internal validity, but increases external validity as greater realism supports generalization outside the strict bounds of the experiment.

---

[22] See Boruch, Robert, (Special Editor). (2005, May). Place randomized trials: Experimental tests of public policy. *The Annals of the American Academy of Political and Social Sciences, 599.*

3.4 Salience and coherence

Salience and coherence are also important aspects of economic experimentation especially in the field setting.

- *Salience* ensures that participants see the experiment and the options offered as personally relevant. If people treat the options with indifference, then they will not select outcomes that matter to them and the entire experiment will fail. A further requirement for saliency is that participants represent the subjects of the policy. A common weakness for much of experimental psychology is that basic ideas were and are validated using undergraduate students. Designing a policy for seniors clearly requires participation by seniors, with experimental options that this group finds personally relevant, to encourage focused participation in the exercise and support generalization to the entire target client group.

- *Coherence* means that the alternatives offered to subjects represent a reasonably complete set for the policy. Participants need to accept that the experiment presents an acceptable representation of reality. In many cases, to ensure logistical feasibility, the researcher must delimit the feasible range of options.

3.5 Financial incentives

Economic experimentalists have always supported the idea of using financial rewards as a way to help participants focus and to increase the salience of the test. Several basic precepts help justify the use of participant payment in experiments.[23]

- *Nonsatiation* means that subjects will always choose options with the greatest personal gain (not necessarily financial).
- *Saliency* means that alternatives are important and relevant to the subject's well-being or in another formulation — rewards must increase the "good" and decrease the "bad."
- *Dominance* means that the rewards of participation outweigh any costs (real and subjective) of participation.
- *Privacy* means that the rewards are private to the individual.

The issue of financial incentives for participation has caused extensive controversy in the literature. Some argue that it distorts preferences, and that rewards from participation contaminate decision-making in experiments. The emerging consensus appears to be that incentives are not only acceptable, but that experiments need financial rewards to ensure focus and realism. This goes beyond the perfunctory attendance fees for a focus group, to even paying cash in relation to success in playing the "game."[24]

3.6 Summary on experiments in economics

This section has very briefly touched on some of the important themes underlying economic experiments, in particular for laboratory and field experiments.

---

[23] See Guala (2005), pp. 232–233.

[24] See Binmore (1999).

4.0 Case studies in laboratory and field experiments

This section presents a series of experimental case studies from the literature. Each case study discusses the nature of the experiment and its possible policy relevance. For the purposes of this study, policy relevance refers to both direct experimental applicability and the application of information, gathered through experimentation, to policy formation. In general, the cases move across the spectrum of laboratory and field experiments, starting with an example of what might be termed a pure laboratory experiment. The studies then present research that moves progressively along the spectrum to field studies.

4.1 Individual motivation

Kritikos and Bolle (2004) identify a history of challenging the self-interested maximizing individual assumption. In fact, many theoretical, empirical studies and experiments seem to support a variety of altruistic, reciprocal, and inequality-averse behavioural assumptions. Though it is not difficult to argue that altruism, reciprocity, and inequality aversion play a role in decision-making, it is far more challenging to identify the degree to which these factors affect decision-making or how these effects may be appropriately modelled.

As it is difficult to observe individuals' motivations in practice, experimentation provides an alternative observational setting for individual behaviour. To explore issues of altruism and reciprocity, these researchers used a variety of two-person games to assess the impact of altruism and reciprocity under various distributional settings.

They assumed that in a one-stage game,[25] reciprocity could not affect individual decision-making, as there was no room for retaliatory or cooperative behaviour motivated by previous actions. On the other hand, in a repeated game setting, both altruism and reciprocity may play a role in players' decisions, and netting out the effect of reciprocity requires comparing the results of one stage and repeated games. By varying the final payout distributions during game play, Kritikos and Bolle further explored whether inequality aversion also affected decision-making. Inequality aversion simply refers to the desire by game participants to see fair outcomes.

> ***The experiment: Approaching fair behaviour — Distributional and reciprocal preferences***
>
> Kritikos and Bolle employed a series of one-stage and multi-stage two-person games where individual decisions dictated the degree of equality and level of personal gain among payoffs. The researchers argued that in one-stage games, where there can be no subsequent retaliation, decisions favouring equality over personal gain are motivated only by altruism. However, in multi-stage games, where selfish decisions may be punished by the other player in subsequent rounds, decisions favouring equality over personal gain are motivated by altruism and feelings of reciprocity.

The results of their *laboratory* experiments are mixed, but tend to show that despite resulting in increased inequality, individuals were often willing to make altruistic choices to benefit other game players. Additionally, under a variety of settings, almost half of the experimental subjects

[25] A one-stage game only supports a single decision by all participants, with no option for change after the initial decision.

made reciprocal choices even when they resulted in increased inequality in the other players' favour. They conclude that as a result, it is not possible to accurately model individual decision-making with the inclusion of only one of the three motivational assumptions explored.

By their own admission, Kritikos and Bolle note a dearth of information exists to accurately model individual utility functions (individuals's goals or want satisfaction), and that their own experimentation did not provide sufficient information to develop an accurate and universally applicable utility functions. They do, however, suggest other experimental lines that may provide insight into more suitable theoretical utility functions.

From a policy perspective, accurately modelling individual motivations allows development of policies that provide appropriate motivational incentives to individuals. For example, Labour Market Adjustment Programming that does not sufficiently consider individuals' altruistic motivations with regard to children and family will not be successful in providing return to work and educational incentives. In this case, the well-being of children and family may be a major motivational factor, and incentives to work or train may remain only partially effective in the absence of policies or programs addressing the factors.

4.2 Eliciting time and risk preferences

Harrison, Lau, Rutström, and Sullivan (2005) note that choices (utility functions) have three dimensions — the actual value of the commodity, the time or duration of use, and uncertainty about the product quality/delivery. Although one may model time and risk preferences relating to uncertainty using a set of simple assumptions, experimental methods attempt to gather time and risk preference information directly from individual behaviour. Such knowledge would be important to HRSDC, for example, because these attributes could influence decisions to invest in education, training, job choices, retirement income, life long learning.

In the laboratory experimental setting, one may model time preferences using a simple technique developed by Coller and Williams (1999), which examines subject responses to a series of questions asking whether they prefer $x today or $x + y in one day. In this setting the choice of $x today implies that an individual's discount rate is higher than y%. Perturbations to x, y, and the time provide precise information about discount rates under a variety of circumstances.

One may experimentally gather risk preference information using a technique developed by Holt and Laury (2002), where experimental subjects choose between monetary lotteries, with differing absolute and relative expected values. The choice of lottery among subjects, as certainty about payoffs increases, provides information about individuals' level of risk aversion. Varying the level of monetary prize associated with the lotteries provides information on risk at a variety of payout levels.

The work of Harrison, Lau, Rutström, and Sullivan applied refinements of both techniques and moved from the laboratory to the field setting. Their work randomly sampled 25,000 individuals from the Danish Civil Registration Office, creating from this 25,000, random sub-samples associated with each Danish municipality. The result was an experimental population that closely matched the demographic profile of the general Danish population, from which Harrison et al. solicited participation in a variety of experiments.

The experiment: Eliciting risk and time preferences using field experiments

Harrison, Lau, Rutström, and Sullivan apply a modified version of a well-established laboratory model to a representative sample of the Danish population. Experimental subjects face two types of choice decisions. The first involved decisions between playing low- and high-risk lotteries with varying expected values. The second involved decisions between small monetary payoffs now or larger monetary payoffs in the future. Decisions favouring low-risk lotteries and immediate payments indicate risk aversion and a high level of time discounting.

These classes of experiments show the policy applicability of more precise measures of risk and time preference knowledge, particularly with regard to asset-based strategies designed to generate benefits to users over long time horizons, and with varying degrees of uncertainty. This may include savings and investment programs for low-income individuals, whose risk and time preferences are not likely representative of the broader population, and are consequently not easily modelled using observational studies that cannot vary risk and discounting assumptions. Experimentation with any target population can provide reliable information on risk and time preferences across a spectrum of policy parameters, which offers a cost-effective approach to policy design.

4.3 Short- and long-term time preferences

Although the Harrison, Lau, Rutström, and Sullivan study used experimental methods to elucidate population-specific time preferences, it did not directly examine the association between short- and long-term time preferences. As such, it ignored the possibility of a stable relationship between individual decisions made while facing various time horizons.

To address this relationship experimentally, Eckel, Johnson, and Montmarquette (2005) looked closely at data from an experiment involving the Canadian working poor. Although they noted that the working poor may fail to save for many reasons, including present orientation, risk aversions, high levels of uncertainty, and of course the pressing need to survive, they attempted to measure the relationship between long- and short-term time preferences for this group.

The experiment: Saving decisions of the working poor — Short- and long-term horizons

Eckel, Johnson, and Montmarquette present experimental subjects from the working poor with three types of decision-making problems.

- The first involved choosing between playing high- or low-risk lotteries with varying expected values.
- The second involved deciding between small monetary payoffs immediately and larger payoffs in a number of weeks.
- The third involved choosing between small monetary payoffs immediately and considerably larger payoffs in a number of years.

While the three decision tasks allowed researchers to individually establish risk as well as short- and long-term time preferences among the working poor, comparison of results helped identify a stable relationship between short- and long-term time preferences among the working poor.

After controlling for numerous individual and personal characteristics, the study finds considerable difference in time preferences for various subgroups within the experimental population. They also find that, while personal characteristics influence time preferences, more risk-averse individuals are generally more present oriented, and that in some cases, while highly present oriented in the short-term, individuals may be far less present oriented in the long-term. Further, the research tentatively concludes that long-term and short-term time preferences correlate sufficiently to use short-term preferences alone to predict long-term preferences.

Determining the association between short- and long-term time preferences has a simple yet practical policy application. As the authors note, in an experimental setting, eliciting short-term time preferences is far less costly than eliciting long-term time preferences. In general, incentives for participation, subject payments, and logistical costs associated with short-term time preference experiments are lower, making them attractive from a cost-effectiveness perspective. Using short-term experiments to predict long-term time preferences would then reduce the cost of experimentation and policy design to support the development of long-term savings and educational programs discussed in previous sections.

4.4 Linking behaviour to program outcomes

As Slonim and Bettinger (2005) state, experimental methods provide a convenient means of assessing many program and intervention impacts that are not easily measured using traditional direct empirical measurement. Because many programs, including those that governments run or support, result in impacts long after their intervention periods, impact assessment using traditional means may require significant evaluation periods and may not provide information about programming in a timely manner. This is particularly true of educational innovations that, while implemented early in a student's education, only show impacts much later in life. The value in establishing the existence of intermediate outcomes in a results chain with extended end outcomes is apparent to any evaluator.

Field experimentation may provide insight into immediate program effects on individual characteristics that, when correlated with future outcomes among program participants, may provide insight into long-term program impacts. This logic guides the field experimentation surrounding educational voucher programs performed in the Slonim and Bettinger study. The key idea is that experimentation that measures the effect of educational voucher programs on student confidence establishes an initial condition supporting the long-term economic impacts of the educational voucher system, because confidence correlates with positive economic impacts.

The experiment: The effects of educational vouchers on confidence — A field experiment to assess outcomes of educational policy

The Children's Scholarship Fund of Toledo provides educational vouchers valid for the reimbursement of up to one half of the cost of private school education. Presented to a random selection of applicant children, the voucher allotment actually created a natural experiment by establishing randomly selected control and treatment groups. By comparing controlled confidence test results among lottery winners and losers, the authors established the net effect of the educational vouchers on child confidence.

The main finding was that lottery winners were not significantly different in terms of confidence either ex-post or ex-ante. However, voucher winners were significantly less overconfident than losers were, when comparing across non-African Americans. Additionally, the authors found that increased child age resulted in higher ex-post and ex-ante confidence, and that children from higher-income homes exhibited higher ex-ante confidence.

Numerous national and local programs involve interventions meant to impact clients over periods of years and even decades. For example, the Social Research and Demonstration Corporation's (SRDC) AVID program, predicated on the link between behavioural traits and academic success, attempts to influence long-term academic success and post-secondary participation among children by fostering specific behavioural habits and attitudes early in their educational experience (SRDC, 2005b). Experimental methodologies such as those used by Slonim and Bettinger provide a means of assessing these programs well before participants enter post-secondary education, providing early information regarding program improvement and adjustment.

4.5 Social capital experimentation

Although a prolific area of study, social capital remains a problematic theoretical concept. Some of this difficulty comes from the variety of definitions associated with social capital. Durlauf (2002) identifies at least three separate and often-used definitions of social capital, each of which has different implications for the practical application of the social capital concept. Beyond this, the definition of social capital varies with some authors using a functional definition and others interpreting it as a causal factor in determining community welfare terms. In very general terms, social capital references individuals, communities, and entire societies. At the individual level, social capital refers to the advantages gained by an individual because he/she connects to various personal, professional, and social networks. At the community level, social capital refers to the benefits a community enjoys when it has a network of volunteers and participation in civic processes, etc. Putnam (2000) extended the idea to American society in general and argued that declining voter turnout, engagement in political parties, and reduced participation in voluntary organizations constitute evidence of declining social capital.

The concept of social capital is challenging theoretically. Implementing an empirical program to examine whether it is indeed a causal factor in individual and community welfare has proven most challenging. Durlauf argues that experimentation is fertile ground for the exploration of social capital.

The experiment: On the empirics of social capital

Durlauf outlines the pioneering Robber's Cave experiment where a group of homogeneous boys was separated into two groups, isolated, and asked to engage in a variety of team-building exercises. When reintroduced to each other, experimenters tracked changes in behavioural norms among the boys relative to those established prior to separation. Experimenters also noted the degree to which competition between groups facilitated collective group action.[26]

[26] This game bears a strong resemblance to the situation depicted in the *Lord of the Flies* by William Golding (1954).

Despite the apparent applicability of experimentation, like that found in the Robbers' Cave experiment, few recent works have used experimental frameworks for the exploration of social capital. However, clear policy applications of social capital theory do exist. For example, First Nations National Child Benefit Reinvestments have, as their largest body of programs, cultural initiatives arguably meant to foster community identity and promote positive child outcomes. Increasing the understanding of how investment in social capital enhances child outcomes, as facilitated by specific cultural programs, would allow for the development of National Child Benefit Reinvestment and other social development programs tailored to individual and community needs.

4.6 Experimental program evaluation

Under many social assistance regimes, as the unemployed find work and gain employment income, government reduces transfer payments. Though at low levels governments may not tax employment income heavily or at all, as earnings rise the reduction transfer payments can result in an effective marginal tax rate for individuals that may be as high as 100%. As a result, individuals securing employment are worse off, as after-tax income remains static, yet leisure time falls and work time increases.

Spermann and Strotmann (2005) attempted to circumvent the need for post-program evaluation by using field experimentation. In Mannheim, Germany, they established treatment and comparison groups among the long-term unemployed and differentially applied a targeted negative tax initiative developed by local administrators. Comparisons of each group's probability of finding and securing employment provided a measure of the effectiveness of the initiative, prior to full program implementation. This field experiment effectively becomes a social experiment, illustrating the fine line that exists among these varieties of studies.

The experiment: The targeted negative income tax in Germany

Spermann and Strotmann studied a negative tax incentive regime in Germany, applied differentially to individuals within and across municipalities, to establish the regime's success or failure. The differential application of the regime within comparable regional settings established experimental treatment and control groups for further analysis. Years after application, comparing the re-employment success of individuals under and not under the regime allowed researchers to establish the impacts of the negative tax incentive regime.

The authors find that the success of the program first relies on the effective administration of the negative tax program. As the effectiveness of the program depends on informing the unemployed about the improved incentives to work under the program, program success was predicated on effective communication regarding programming. When administered correctly, they found that there was a positive effect on individuals' probability to participate in the labour market. They further note that the results are in line with international works, including those involving large-scale social experimentation.

For policy, the experimental methods used by Spermann and Strotmann have a dual advantage. First, they provide information on program effectiveness prior to full program implementation. This allows for low-cost, early program adjustments during the development stage rather than high-cost evaluation and program adjustments usually long after program administration and infrastructure are in place. In the case of large-scale government programs, like the Labour

Market Development Agreements, improved targeting could result in significant administrative savings.

Second, as the experimental methodology costs less and is logistically simpler than many alternative impact evaluation techniques, yet provides similar if not more reliable results, it represents a cost-effective and accurate means of evaluating programming.

4.7 Targeted educational support policies

Many see investments in human capital, through education and training, as a means of addressing issues of unemployment and underemployment among the poorest segments of the population. However, programs meant to address these problems must understand the motivation and decision-making processes of their target clientele to be effective. Simply providing educational incentive packages, of varying types, will not address either unemployment or underemployment if uptake, participation and learning are low.

To address various issues surrounding the effectiveness of human capital development interventions, Eckel, Johnson, and Montmarquette (2005) conducted a number of field experiments with a variety of subject groups. The use of a diversity of subjects has allowed Eckel et al. to establish results specific to a variety of income and demographic groups, thereby increasing the external validity of their work.

> ***The experiment: Investment by working poor in human capital***
>
> Eckel, Johnson, and Montmarquette attempt to identify preferences for educational savings among the working poor by soliciting responses to a variety of choice questions regarding current consumption and savings for future education. Experimental subjects are asked to choose between a small immediate monetary payoff and a larger monetary contribution to an educational saving fund meant to support future education or training. Varying the level of immediate and educational saving payoffs allowed experimenters to establish the level of incentive necessary to make individuals among the working poor invest in education.

Methodologically, the researchers present subjects with a series of options regarding educational and human capital investment options and ask that they identify their preferred choice among the options. For example, when examining educational savings decisions among the working poor, they asked if the subject would prefer an immediate cash payment or a larger sum set aside for educational purposes. The random payment of one or more of the subject's choices assured that the subject's decision accurately reflected individual preferences.

Through experimental manipulation, it is possible to draw conclusions about the choice decisions of various groups regarding educational incentives. For example, they noted that grant programs were a universally preferred means of encouraging investment in education and that matching grant programs were universally preferred to loan programs.

This type of information is of particular importance to policy development, as educational incentive and human capital development programs must provide appropriate incentives for their target clients. These experiments demonstrate the potential for individual development accounts among the working poor and served as the foundation for the *learn*$ave experiment, developed

by Social and Enterprise Development Innovations and evaluated by the Social Research and Demonstration Corporation.

4.8 Using choice experiments in estimating the welfare wall

Choice experiments are starting to emerge in social policy design (see List, 2002). Long used in environmental economics as a method to elicit preferences and valuations of public goods such as environmental preservation, these techniques have significant potential policy design. One application uses a discrete choice experiment to estimate the reservation wage of social assistance clients, which is a proxy for estimating the height of the welfare wall.[27]

The reservation wage is a subjective value influenced by many factors such as the nature of the work sought (with dangerous employment and "off-hour" work requiring a higher wage), distance to job, other income earned by the household, family responsibilities, etc.

Most survey-based studies on the reservation wage use a direct question such as, "How much would the wage need to be for you to accept a position now?" This question has a direct parallel with market research, where researchers pose the question "How much are you willing to pay for product 'X'?" Market researchers have recognized that directly posing a pricing question may produce strategic response bias, where respondents purposely under- or overstate their willingness to pay in an attempt to sway results of the research.

The experiment: Discrete choice methods for estimating labour supply.

On a client survey conducted as part of the National Child Benefit evaluations after covering the questions pertaining to satisfaction and other aspects of the evaluation, respondents were randomly assigned to two groups. Respondents in each group agreed or disagreed to a specific question: "Would you accept a job at $Y per hour?" where Y varies with the group. Those who respond "YES" can be asked whether they would accept a job at a test wage of $Y – $1, and those who reply "NO" may be tested with $Y + $1. The percentage of respondents who answer YES (i.e., they would accept a job that provides a "YES") may be plotted against the three test wages (Y, Y – $1, Y + $1). With two groups, a total of 6 data points results, which when multiplied by the sample size of the survey and combined with the covariates collected on the respondent provides a very rich database.

Reservation wage
(Computed as the test wage accepted or the reservation wage offered if both test wages rejected)

Group	SA*	Non-SA	Total
	(Sample size in parentheses)		
All clients	$9.97 *(711)*	$10.67 *(1,011)*	$10.39 *(1,728)*
Single parents	$10.04 *(509)*	$10.77 *(302)*	$10.33 *(814)*
Dual parents	$9.79 *(202)*	$10.63 *(709)*	$10.44 *(914)*

* Receipt of SA in the last year (prior to interview date)

[27] This work, originally completed by PRA Inc. as a technical study, supported the evaluation of the National Child Benefit.

This research reveals a reservation wage that exceeds the minimum wage (as of 2001) and also serves as a direct measure of the welfare wall. For the most part, these results emerging from a discrete choice experiment appear reasonable, but it is important to note that the survey collected respondents' reported willingness to accept a job at a stated wage. This is not the same as observing whether a respondent accepts a job if it is really offered at that wage. This "artificiality" is an important barrier to wider acceptance of these experimental approaches. Critics view experimental data as fundamentally unreal.

4.9 Using stated choice experiments to design an income support package

Stated choice methods alternatively referred to as *conjoint analysis* methods, map preferences across a range of products or services so that individual features (attributes) may be studied both independently and simultaneously. Harrison (2006) is a recent example of how these methods may be applied to social policy.

The term conjoint is a contraction of consider jointly.[28] The application of conjoint experiments to the design of an income support program requires two essential steps:

- First, the research must deconstruct the program/policy elements into *attributes* and *levels*, as shown in the example below. This can be challenging enough for a commercial or consumer good or service, let alone a social or economic policy directed to a specific target audience.

- An important feature of conjoint studies is their ability to support experimental variation in attributes and levels. However, as the number of levels for each attribute increases, the number of different options that participants in the experiment need to value rises very sharply.[29] Fractional factorial designs[30] present a subset of all possible packages to each participant to support considered decision-making.

---

28 A recent reference for these methods is Louviere, J., Hensher D., & Swait J. (2000). *Stated choice methods: Analysis and application.* Cambridge: Cambridge University Press.

29 Consider an ordinary consumer service such as a restaurant. If one selects three attributes, each with three levels, the number of different packages is $3^3 = 27$, an onerous exercise. Fractional factorial design presents a subset of all choices to each participant so that across all participants in a series of focus groups all packages are evaluated several times.

30 A factorial experiment is an experiment whose design consists of two or more factors, each with discrete possible values or "levels", and whose experimental units take on all possible combinations of these levels across all such factors. Such an experiment allows studying the effect of each factor on the response variable, as well as the effects of interactions between factors on the response variable. **Fractional factorial designs** are experimental designs consisting of a carefully chosen subset (fraction) of the experimental runs of a full factorial design.

The experiment: Stated choice experiment in social policy preferences

As a technical appendix to the evaluation of the National Child Benefit (NCB), focus group participants evaluated various options for income and in-kind support. The range of choices presented to participants appears below. This is a stated choice experiment with 5 attributes, with 2 or 3 levels. Participants rated policy packages such as "no cash benefits, free child care, food voucher worth $50, no clothing voucher, and free after-school club."

Attributes and levels for the conjoint experiment — Social assistance package			
Attributes	**Levels**		
Additional monthly cash benefit	No cash benefit	$50 per month	$75 per month
Additional child care while working	$2.50 cost per child per day	Free child care	-
Voucher for food	No voucher	$50 per month	$75 per month
Voucher for clothing	No voucher	$75 per month	-
After-school club (recreation and homework help for children)	$1.50 per child per day cost	Free after-school club	-

A stated choice experiment consists of a trial or a series of trials, typically conducted as part of a focus group, in which the researcher makes purposeful changes to the attributes/levels of a package in order to observe and identify possible explanations for changes in the response variable.[31] In the simplest form of analysis, the resulting data set is a regression model, where the valuations of packages become the dependent variable and the packages are coded using dummy variables.

This experiment revealed that this group (National Child Benefit Supplement recipients) valued cash support much higher than in-kind support and did not value child care to assist participation in the workforce. This latter finding is at odds with the goals of the NCB; however, it aligns exactly with other research conducted as part of the evaluation.

5.0 Summary

5.1 Final synthesis: How laboratory, field, and social experiments differ

The terms laboratory, field, and social experiments tend to overlap and converge and while the differences tend to be matters of degree rather than kind, the boundaries remain somewhat fuzzy. Thus, proponents of one approach can either draw extreme distinctions, or minimize the differences to promote their favoured technique. Table 1 attempts to summarize these essential distinctions.

[31] This is a brief summary of the ideas underlying stated choice modelling. See Mason (2006) and references therein for a more complete explanation.

Table 1: Laboratory, field and social experiments in economics compared			
	Laboratory experiment	**Field experiment**	**Social experiment**
General feature	Most lab experiments are completed using paper exercises in a lab, on-line, or using pen and paper exercised in a highly controlled setting.	Many field experiments can be grafted onto other research methods, such as focus groups and surveys.	Social experiments may replicate elements the policy exactly for the treatment/program, group.
Subjects	Non-standard populations (typically university students).	Participants are typically selected from the target population to increase generalizability (external validity).	Participants are selected as if a policy were in place. The control group is exactly eligible for participation.
Subject randomization	Limited randomization of participants; experimental protocols enforce the randomization of policy options among the participant set.	A mix of participant randomization and protocol randomization.	Participants randomized into treatment and control groups.
Compensation	Small financial incentives as a reward for winning games.	Typically, none aside from compensation to participate in the survey or focus group; financial incentives can be used.	Participants are provided financial and other compensation as if the policy (e.g., earnings supplement) were in place.
Policy/theory testing	Every aspect of the policy experiment is carefully controlled to increase internal validity. This supports rapid testing of a range of policy options.	Policy testing range is good and only limited by the sample sizes.	Policy testing is limited to a few features (typically less than five) that can be supported by treatment and control groups. Every policy dimension requires a significant increase in sample sizes.
Design time and cost	Low, conditioned only by the time needed to program new features into the computer "game." Test run in accelerated time and allows for inexpensive testing of hypotheses/concepts.	Moderate, conditioned by time for experimental design time, enrolment and execution of data collection (survey and focus groups).	High because of the logistics of designing and implementing an exact replica of a policy context.
Internal validity	High – the games are highly controlled; thereby preserving the experimental policy control.	Moderate – data collection context introduces variability in policy control.	Low – attrition (people leaving the experiment) and logistical complexity threaten to upset policy control.
External validity (generalization to population)	Low – e.g., university students may not be representative; however, in some cases, they may be. Or, laboratory conditions do not simulate field situations.	Medium – survey and focus groups can be enrolled to more closely resemble the target policy group.	High – participation exactly replicates the intended target policy group (e.g., single parents on social assistance).
Significant strengths	Low cost and rapid policy/theory testing and validation. Results within weeks.	Moderate cost, especially when grafted onto existing data collection. Long record of success in market research. Results within months.	High credibility of results since the policy context and subjects are an exact match to the "real" world.
Significant weakness	Low external validity since subjects are non-standard. Application to real world is problematic and context is limited by the lab setting.	Salience is limited by the subjects capacity to "imagine" the policy context in the context of the experiment (telephone survey, focus group).	Attrition and panel conditioning (subjects become used to the repeated follow-up interviewing) upset the integrity of the experiment. High cost and delayed results (at least 1 year after the participation).

5.2 Conclusion

Laboratory, field, and social experiments can contribute important information to policy planners because they have much to reveal about basic human behaviour and the decision-making process. They can also reveal important motivations and conditioning variables for these motivations, which influence decisions to invest both money and time. These are important considerations for the design of asset-based policies and the creation of incentives for participation in training programs because, despite having spent millions on evaluations of these types of programs in Canada, there is little information on what motivates an EI client or SA recipient to participate in training, seek and maintain employment, etc.

The main impediment to application of laboratory experiments is that they lack external validity. Field experiments relax the strict controls imposed in the laboratory in an attempt to increase the generalizability to wider populations. A simple extension of the laboratory experiment into the field may involve the inclusion of a wider range of participants. Collecting personal and socio-economic information on the increasingly diverse experimental participants allows the researcher to use the response data in multivariate models and apply results to programs targeting specific populations.

As the experiment moves further into the field, the strict controls of the laboratory fade, and researchers need to find other methods to ensure that sufficient internal validity exists in order to come to valid conclusions about human motivation and behaviour. Constructed field experiments, such as the study on education vouchers, retain significant internal validity through the randomization of participants. However, discrete choice and stated choice (conjoint) methods often benefit from the inclusion of a range of other covariates in the context of a multivariate model. In these cases, the researcher introduces statistical control after the experiment to augment the usefulness of the experimental results.

As a final note, it is important to recognize that the data collection and multivariate analysis needed to study many policy issues are exceedingly costly. Smaller-scale field experiments offer important logistical and cost advantages in studying these questions. For example, qualitative research surrounding Employment Benefits and Support Measures associated with Labour Market Development Agreements reveals that the experience an EI client has with an employment centre, and especially with his or her counsellor, is an important determinant to successful retraining. Small-scale experimentation may address a range of hypotheses regarding the organization and even “style” of counselling, which would be prohibitively costly using traditional multivariate methods.

HRSDC may benefit from experimenting with such smaller-scale experiments.

Bibliography

Amiel, Y., Cowell, F., & Slottje, D. (2004). Why do people violate the transfer principle? Evidence from educational sample surveys. In F. Cowell (Ed.), *Inequality, welfare and income distribution: Experimental approaches* (pp. 1–16). Oxford: Elsevier Ltd.

Beckman, S.R., Formby, J.P., & Smith, W.J. (2004). Efficiency, equity and democracy: Experimental evidence on Okun's leaky bucket. In F. Cowell (Ed.), *Inequality, welfare and income distribution: Experimental approaches* (pp. 17–42). Oxford: Elsevier Ltd.

Bernasconi, M., & Dardanoni, V. (2004). An experimental analysis of social mobility comparisons. In F. Cowell (Ed.), *Inequality, welfare and income distribution: Experimental approaches* (pp. 55–84). Oxford: Elsevier Ltd.

Binmore, Y. (1999). Why experiment in economics? *Economic Journal, 109*, F16–24.

Boruch, Robert, (Special Editor). (2005, May) Place randomized trials: Experimental tests of public policy. *The Annals of the American Academy of Political and Social Sciences, 599*.

Bosmans, K., & Schokkaert, E. (2004). Social welfare, the veil of ignorance and purely individual risk: An empirical examination. In F. Cowell (Ed.), *Inequality, welfare and income distribution: Experimental approaches* (pp. 85–114). Oxford: Elsevier Ltd.

Botelho, A., Harrison, G.W., Hirsch, M.A., & Rutström, E.E. (2005). Bargaining behavior, demographics and nationality: What can the experimental evidence show? In J.P. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics, Vol. 10. Research in experimental economics* (pp. 337–372). Oxford: Elsevier Ltd.

Burtless, G. (1995). The case for randomized field trials in economics. *Journal of Economic Perspectives, 9(2),* and Heckman, J., & Smith, J. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives.*

Card, D., & Krueger, A.B. (1995). *Myth and measurement: The new economics of the minimum wage*. Princeton: Princeton University Press.

Cardenas, J.C., & Carpenter, J.P. (2005). Three themes on field experiments and economic development. In J.P. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics, Vol. 10. Research in experimental economics* (pp. 71–123). Oxford: Elsevier Ltd.

Carpenter, J.P., Burks, S., & Verhoogen, E. (2005). Comparing students to workers: The effects of social framing on behavior in distribution games. In J.P. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics, Vol. 10. Research in experimental economics* (pp. 261–289). Oxford: Elsevier Ltd.

Carpenter, J.P., Harrison, G.W., & List, J.A. (2005). Field experiments in economics: An introduction. In J.P. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics, Vol. 10. Research in experimental economics* (pp. 1–15). Oxford: Elsevier Ltd.

Chamberlin, E. (1948). An experimental imperfect market. *Journal of Political Economy, 56,* 95–108.

Checchi, D., & Filippin, A. (2004). An experimental study of the Poum hypothesis. In F. Cowell (Ed.), *Inequality, welfare and income distribution: Experimental approaches* (pp. 115–136). Oxford: Elsevier Ltd.

Coller, M., & Williams, M.B. (1999). Eliciting individual discount rates. *Experimental Economics, 2*, 107–127.

Cummings, Ronald G., Holt, Charles A., and Laury, Susan K. Using laboratory experiments for policy making: An example from the Georgia irrigation reduction auction. Department of Economics. Andrew Young School of Policy Studies, Georgia State University. http://www2.gsu.edu/~wwwenv/waterPDF/W2002003.pdf

Davidovitz, L., & Kroll, Y. (2004). On the attitude towards inequality. In F. Cowell (Ed.), *Inequality, welfare and income distribution: Experimental approaches* (pp. 137–147). Oxford: Elsevier Ltd.

Davis, D., & Holt, C. (1993). *Experimental economics.* Princeton: Princeton University Press.

Durlauf, Steven N. (2002). On the empirics of social capital. *The Economic Journal, 112*, F459–F479.

Eckel, C., Johnson, C., & Montmarquette, C. (2005). Saving decisions of the working poor: Short- and long-term horizons. In J.P. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics, Vol. 10. Research in experimental economics* (pp. 219–260). Oxford: Elsevier Ltd.

Eckel, Catherine, Johnson, Cathleen, & Montmarquette, Claude. (2002). *Will the working poor invest in human capital? A laboratory experiment.* Social Research and Demonstration Corporation. Working Paper 02-01.

Eckel, Catherine, Johnson, Cathleen, & Montmarquette, Claude. (2003). *Fostering adult education: A laboratory experiment on the efficient use of loans, grants and saving incentives.* Social Research and Demonstration Corporation. SRDC Working Paper series 03-09.

Falk, Armin, & Fehr, Ernest. (2003) Why labour market experiments? *Labour Economics, 10,* 399–406.

Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists.* Cambridge: Cambridge University Press.

Gendler, T.S. (2000). *Thought experiment: On the powers and limits of imaginary cases.* New York: Garland and at http://en.wikipedia.org/wiki/Thought_experiment (although Wikipedia changes almost weekly).

Giles, Jim. (2006, July). The trouble with replication. *Nature, 442,* 344–347.

Golding, William. (1954). *Lord of the Flies.* London: Faber and Faber.

Guala, F. (2005). *The methodology of experimental economics*. New York: Cambridge University Press.

Harrison, G.W. (2005). Field experiments and control. In J.P. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics, Vol. 10. Research in experimental economics* (pp. 17–50). Oxford: Elsevier Ltd.

Harrison, G.W., & List, J.A. (2004, December). Field experiments. *Journal of Economic Literature, 42(4),* 1013–1059.

Harrison, G.W., Lau, M.I., Rutström, E.E., & Sullivan, M.B. (2005) Eliciting risk and time preferences using field experiments: Some methodological issues. In J.P. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics, Vol. 10. Research in experimental economics* (pp. 125–218). Oxford: Elsevier Ltd.

Heckman, J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *Quarterly Journal of Economics, 115*, 45–97.

Heckman, J., Ichimura, H., & Todd. P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies, 65*, 261–294.

Holt, C.A., & Laury, S.K. (2002). Risk aversion and incentive effects. *American Economic Review, 92(5),* 1644–1655.

Isaac, R. (2001). *Research in experimental economics, Vol. 8*. Amsterdam: Elsevier Science; Greenwich, CT: JAI Press Inc.

Kagel, J.H., & Roth, A. (1995). *The handbook of experimental economics.* Princeton: Princeton University Press.

Kritikos, Alexander, & Bolle, Friedel. (2004). Approaching fair behavior: Distributional and reciprocal preferences. In F. Cowell (Ed.), *Inequality, welfare and income distribution: Experimental approaches* (pp. 149-180). Oxford: Elsevier Ltd.

Lehrer, Jonah. (2006, October). Driven to market. *Nature, 443*, 502–504.

List, John A. (2002). Using choice experiments to value non-market goods and services. (Working Paper).

Louviere, J., Hensher D., & Swait J. (2000). *Stated choice methods: Analysis and application.* Cambridge: Cambridge University Press.

Mason, Greg. (2006). Applying choice experiments to the design of social programs. (Working Paper). Winnipeg, MB.

Mason, Greg. (2006). The welfare wall, reservation wages, and labour force attachment. (Working Paper). Winnipeg, MB.

Oreopoulos, Philip, (2006, March). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review, 96(1),* 152–177.

Ortmann, A. (2005). Field experiments in economics: Some methodological caveats. In J.P. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics, Vol. 10. Research in experimental economics* (pp. 51–70). Oxford: Elsevier Ltd.

Plott, C.R. (1987). Dimensions of parallelism: Some policy applications of experimental methods. In A.E. Roth (Ed.), *Laboratory experimentation in economics* (pp. 193–219). New York: Cambridge University Press.

Putnam, R. (2000). *Bowling alone.* New York: Simon and Schuster.

Sherif, M., Harvey, O., Hood, W., & Sherif, C. (1961). *Intergroup conflict and cooperation: The Robbers' Cave Experiment*. Norman: Institute of Group Relations, University of Oklahoma, reprinted by Wesleyan University Press, 1998.

Slonim, R., & Bettinger, E. (2005). The effects of educational vouchers on confidence: A field experiment to assess outcomes of educational policy. In J.P. Carpenter, G.W. Harrison, & J.A. List (Eds.), *Field experiments in economics, Vol. 10. Research in experimental economics* (pp. 291–335). Oxford: Elsevier Ltd.

Smith, V.L. (1994). Economics in the laboratory. *Journal of Economic Perspectives, 8* (Winter), 113–131.

Spermann, Alexander, & Strotmann, Harald. (2005). *The targeted negative income tax (TNIT) in Germany: Evidence from field experiments*. Bonn: Institute for the Study of Labour (IZA), University of Freiburg.

SRDC. (2006, June), *Implementing social experiments*. Mimeo.

SRDC. (2005). Design and implementation of a program to help the poor save: The *learn*$ave project. Retrieved August 21, 2006, from http://www.srdc.org/en_what_we_do.asp?id=1436

SRDC. (2005b). Social Research and Development Corporation website. Accessed September 5, 2005, from http://www.srdc.org/en_what_we_do.asp?id=1436

Thaler, R. (1987). The psychology of choice and the assumptions of economics. In A.E. Roth (Ed.), *Laboratory experimentation in economics* (pp. 99–130). New York: Cambridge University Press.

Tucker, Al W. (1950). *A two-person dilemma*. Mimeo. Stanford University. Published under the heading “On jargon: The prisoner’s dilemma.” *UMAP Journal*, *1* (1980), 101.

von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behaviour.* Princeton: Princeton University Press.