

Experimental Design Notes*

1 Introduction

We have previously discussed how microeconomic systems can be implemented as economic experiments. I have also provided some slides as a sample of how a particular economic experiment (the one on exit) was developed. This set of notes is concerned with formalizing some of the concepts discussed in that example.

Recall that there should be a reason for all decisions made by the experimenter when designing the experiment. Some design decisions will be dictated by the proposed question, some by the budget, some by previous research, and others because there is only one option.¹ Based on the chosen topic, there should be focus variables, which are the variables of interest for the particular topic. These variables are the ones the experimenter deems important enough to try to understand the effect that changing the variables has on outcomes. There are also other factors that may affect the results of the experiment, and these are nuisance variables. Both types of variables are present in economic experiments, and recall that the goal is to isolate the effect of the focus variables and to minimize any effect the nuisance variables may have on the results. How is this done? By obtaining independence through control and randomization.

2 Control

All experimenters have the ability to set many variables. Recalling the exit experiment, there were two focus variables, the cost distribution (was it low uncertainty (LU) or high uncertainty (HU)) and the framing of the experiment as an opportunity cost (OC) or fixed cost (FC) experiment. There were many other variables (# of rounds, # of subjects, initial endowment, etc.) that were not focus variables, although they could (and maybe should) have been. Some variables, such as the # of rounds and the endowment in the FC sessions, were held constant across sessions, while others were set at different “values” (like the cost distribution). In the exit experiment there were:

- 2 sessions of 12 subjects with HU and OC
- 2 sessions of 12 subjects with HU and FC
- 2 sessions of 12 subjects with LU and OC
- 1 session of 12 subjects and 1 session of 10 subjects with LU and FC

Consider the HU sessions. There were 12 subjects and 20 rounds in each session, so for each session there would be 480 values drawn from the cost distribution of $[40, 460]$ and 1920 values needed for all 4 sessions. What did we actually do? We drew 480 values, and for each subject/round pair in the HU sessions we held the value constant. The best way to see this may be to look at it in an abbreviated table. Let HU_OC_1 be the first session of HU and OC; HU_OC_2 be the second session of HU and OC; HU_FC_1 be the first session of HU and FC; and HU_FC_2 be the second session of HU and FC:

*Based on Chapter 4 of Friedman and Cassar. Also, the primary example is from the PowerPoint slides on developing the exit experiment.

¹There are probably other categories that can be defined.

Subject	Round	HU_OC_1	HU_OC_2	HU_FC_1	HU_FC_2
1	1	$v_{1,1}$	$v_{1,1}$	$v_{1,1}$	$v_{1,1}$
1	2	$v_{1,2}$	$v_{1,2}$	$v_{1,2}$	$v_{1,2}$
2	1	$v_{2,1}$	$v_{2,1}$	$v_{2,1}$	$v_{2,1}$
2	2	$v_{2,2}$	$v_{2,2}$	$v_{2,2}$	$v_{2,2}$
20	5	$v_{20,5}$	$v_{20,5}$	$v_{20,5}$	$v_{20,5}$

The same process was used for the LU OC and FC sessions. Obviously the values could not be held constant across LU and HU treatments since the cost distribution changes.

Another "unseen" variable is how subjects were matched in the experiment. While subjects never knew who the other subject in their pair was, it is possible that if subject 1 in round 1 had a cost of 250 and played subject 12 with cost of 402 in HU_OC_1 and played subject 3 with cost of 97 in HU_OC_2 that this random pairing might affect the behavior of subject 1 (and subjects 3 and 12 as well) for the rest of the experiment. We wanted to minimize these effects so the subject pairs were kept constant across all HU treatments – if subject 1 had value $v_{1,1}$ in round 1 of HU_OC_1 and played subject 6 who had value $v_{6,1}$, then this was also true in HU_OC_2, HU_FC_1, and HU_FC_2. In the LU experiments this was true for both OC sessions and the FC session with 12 subjects. While the subjects in the LU_FC session with 10 subjects did have values consistent with the other LU sessions, since there were only 10 subjects the pairs could not be held constant. All of this was done to minimize nuisances. If the results of the experiment were vastly different across these sessions (especially if the results between the two sessions with the same treatment variables were different) when we held these things constant, then it might suggest that the behavior in the experiment is highly reflective of some other uncontrolled nuisance variable. If we began by NOT holding the values and subject matching constant across the sessions and obtained vastly different outcomes, then it could have been due to the order in which the subjects saw the costs or the outcomes that occurred due to the different subject matchings. On the other hand, if the basic results (high cost firms exited before low cost firms, and as the difference in cost between the two firms increased efficient exit occurred with a higher probability) held as we mixed one (or both) of those 2 items (the realized cost draws and the subject matching) then we would have obtained observations on more distinct pairs of costs.

2.1 Friedman and Cassar on Control

1. Control all the variables you can
2. Control your focus variables as treatments (this is OC vs. FC and LU vs. HU)
3. For most treatments two levels are sufficient to detect their effects
4. Separate the levels widely so the effects will be evident (the discussion about the [40, 460] and the [95, 405] ranges for the cost distributions in the exit experiment)
5. Most nuisances should be controlled as constants
6. Nuisances that you might think interact with a focus variable should be considered as treatments (remember that an experimenter cannot do everything, so choose the treatment variables that will provide the most interesting story, and then if time and money permit consider extending the experiment or testing the boundary)
7. Vary treatments independently

3 Independence

Oftentimes two variables have a dependent relationship, where knowing one variable provides some information about the other variable. Two variables are independent if knowledge of one variable provides no information as to the value of the other.² The text discusses an example of “What makes plants grow better

²Somewhere along the way you likely learned that independent variables have a covariance of 0, although a covariance of 0 does not necessarily imply that two variables are independent.

under trees?" and provides two potential causes: bird droppings and shade. Apparently the two are highly correlated. One potential method of determining the cause is to isolate each variable. Thus, a plot of land could be divided so that $\frac{1}{4}$ receives droppings and shade, $\frac{1}{4}$ receives only shade, $\frac{1}{4}$ receives only droppings, and $\frac{1}{4}$ receives neither. The key is to make sure any potential nuisances (soil quality, weather, etc.) are minimized or kept as constant as possible across the 4 plots of land.

4 Randomization

Randomization can be used to control for problems that arise from nuisance variables and to gain independence. If treatment conditions are assigned randomly, then with enough trials eventually the treatments will be independent of all uncontrolled variables. There are of course two major obstacles to gaining independence – time and money. In order to run the number of treatments necessary to gain "true" independence, much of both is needed. The question then becomes what types of experimental designs can minimize the effects of uncontrolled variables given time and money constraints.

4.1 Experimental Designs

Using the exit experiment as an example, there were two cost distributions and two frames. This 2x2 design creates 4 potential cells:

		Frame	
		OC	FC
Uncertainty	HU	HU-OC	HU-FC
	LU	LU-OC	LU-FC

The following methods can be used to reduce nuisances. All methods have their own benefits and drawbacks.

1. **Completely randomized design** – In the exit experiment there are 4 treatments. The completely randomized design would assign an equal probability (in this case 25%) that any of the 4 treatments is randomly drawn from an urn. The treatment drawn is then run. The "ball" representing the treatment is placed back in the urn, and a new draw is made. The new treatment drawn is then run. Repeat this as many times as possible. Hopefully one can see how this random drawing of the treatments should eventually lead to observations independent of all uncontrolled variables. While this is the best method for gaining independence, there are practical problems with this method. One is that the experimenter may end up with more observations in some treatment cells than others. The worst case scenario is that the experimenter has a bad run of luck and never draws the "ball" for one of the treatments, or at least never draws it while his budget is positive.
2. **Factorial** – In the exit experiment there are 4 treatments, each of which was run twice, for a total of 8 sessions. The factorial design is similar to the completely randomized design, except that now once a "ball" representing a treatment is drawn that ball is NOT replaced. The drawback is that now the experimenter may still get some interdependence between sessions just by luck of the draw. The benefit is that the number of sessions is prespecified, and the experimenter will have observations in all treatment cells.
3. **Fractional factorial** – When there are a large number of treatments, only a fraction of them are actually tested. When beginning experimental work it would be better to focus on experiments with a small number of treatment variables so that fractional factorial design can be avoided as judgement calls are needed to pare down the number of treatments.
4. **Crossover** – With a crossover design more than one treatment is run in one session. These are also called ABBA and BAAB designs, or ABA and BAB designs, where A and B refer to two different treatments. With the exit experiment, it could have been designed so that in the first 5 periods the subjects saw the LU distribution, in the next 10 periods they saw the HU distribution, and in the last 5 treatments they saw the LU distribution again. This design choice will minimize differences due to specific individual characteristics since all individuals go through multiple treatments. One

must be careful to use different orders (both ABBA and BAAB) to ensure that the order in which the individuals see the treatments is not affecting the results (or if it is affecting the results how it affects them).

5. **Within-subjects and between-subjects** – In a within-subjects design each subject sees all levels of a treatment variable. This is similar to the crossover design. In a between-subjects design each subject sees just one level of a treatment, but different subjects see different levels. The within-subjects design will allow for intrapersonal comparisons as subjects see more than one treatment. The between-subjects design facilitates familiarity with the design and minimizes confusion that may arise from the design.
6. **Matched pairs** – In a matched pairs design subjects would make decisions simultaneously for multiple treatments. For instance, in the exit experiment subjects might be in the LU treatment and be asked to make a decision as to when to exit if they were in an OC environment or an FC environment. Their decisions would be recorded and either FC or OC would be randomly chosen for actual play of that round. Again, this design allows the experimenter to see if the same subject responds differently to the different treatments.
7. **Baseline neighborhood** – In this design a baseline experiment is conducted. After the baseline is conducted one treatment is added and then a new session(s) is run using the baseline and the one additional treatment. After those sessions are run a second treatment is added and new sessions are run using the baseline, first treatment and second treatment. Essentially, the experimenter builds up to the final treatment which contains all the treatments, but does not run all possible combinations of treatments (for example, there would be no baseline plus second treatment variable session run unless the first treatment variable was also included).

5 Potential Nuisances

The bulk of the list is from Friedman and Cassar but I have added a few others.

1. **Learning** – Behavior may change as the session progresses as understanding of the game being played deepens. One way to control for learning effects is to use a balanced crossover design, such as ABBA and BAAB. In the ABBA session, if behavior in the first A treatment and the last A treatment is different, then learning might be present. Another option is to ignore the first few rounds of data when performing data analysis.
2. **Experience** – Essentially this is learning across sessions. As a general rule, economic experiments do not use subjects who have already participated in an experiment of a particular type. As an example, if a subject participated in the HU-OC treatment of the exit experiment that subject was not allowed to participate in the HU-FC, LU-OC, or LU-FC treatments nor was the subject allowed to participate in any other sessions of HU-OC. In some established laboratories it may be possible to exclude subjects who have participated in broader categories of experiments. For example, if one researcher has run an experiment on 1st-price sealed bid auctions, a second researcher may be able to exclude that same subject from her experiment on 2nd-price sealed bid auctions.
3. **Boredom and fatigue** – It is best to get data from subjects who are "paying attention". If an experiment is too long subjects may become bored or tired and dominance (remember Smith's precepts) may be lost. Typically experiments are kept under 2 hours, and this includes signing subjects in, reading instructions, execution of the experiment, and payment of subjects. Many experiments (particularly one-shot experiments) are shorter, and in some cases it may be necessary to continue the experiment over multiple days. But it is best to minimize boredom and fatigue in the experiment.³

³Some experimenters are of the opinion that subjects must be constantly working on some task, or that down time should be very short. The exact definition of "very short" is debateable, and recall that early experiments were hand-run and there may have been considerable down time in these experiments.

4. **Extracurricular activity** – Unless communication is a treatment variable it is best to keep subjects from communicating from one another. There is a quote from Adam Smith about businessman getting together for innocent conversation and ultimately the conversation turns to business. If the goal is to see how the institution or environment affects behavior of an individual and not how behavior is affected by other individuals then communication should be disallowed. On the other hand, with the rise of chat rooms and text messaging it is possible to incorporate communication into an experiment through one of these methods. This also allows the experimenter to record the communication that takes place, which may provide some interesting insights.
5. **Self-selection** – People are going to participate in those activities in which they are interested. If the experiment is on economic decision-making, this may appeal to subjects who are interested in economics. If the experiment was biological decision-making, then it may appeal to subjects who are interested in biology. Hopefully the cash payment will appeal to a broad class of subjects, but care needs to be taken to make sure the sample is representative of the population. One thing that experimenters try to do when announcing an experiment is to NOT disclose the particulars of the experiment, but simply tell the subjects when and where the experiment will take place. If an experimenter announced that an auction experiment was going to be run then maybe people who liked auctions would be more likely to participate and if the experimenter announced that an experiment on financial markets was going to be run then perhaps people interested in finance would be more likely to participate. To be honest, I'm not sure this is a bad thing in some types of experiments, and there has been a recent rise in field experiments that use subjects who are familiar with a particular setting.
6. **Experimenter effects** – It is possible that the experimenter himself or herself can affect the experiment, and this effect may be subtle. Sometimes an experimenter has a bad day, and the experimenter's mood may be transferred to the subjects. This is one reason why experimental economists use a pre-set script of instructions, so that the experimenter doesn't unduly influence subjects using different language from session to session. It is also possible that two otherwise identical experiments run by two different experimenters consistently get different effects. It may be that some subjects respond differently to the experimenter based on personal characteristics of the experimenter such as gender, race, age, etc. These effects have not been studied in that much detail, and in experiments in the hard sciences as long as the experimenters use the same protocol the hydrogen atom doesn't care about the personal characteristics of the experimenter. So it may be best to have the same person run all the sessions of the experiment simply to remove potential experimenter effects.
7. **Screen and experimental design effects** – Again, consider the exit experiment. Instead of using a real-time process where subjects observe their round profits increasing or decreasing it would have been possible to design the experiment so that subjects simply made a choice of what time to exit conditional on whether or not the other subject chose to exit before that time. This is essentially the same question without the real-time component of the experiment. The question to ask is whether or not subjects behave differently in these two types of environments and, if so, which is more representative of reality. This non-real-time design does have some benefits, as it would provide more data than the real-time experiment. In the next section we will discuss data analysis.