CAMBRIDGE

Cambridge University Press & Assessment
978-1-009-50035-7 — Advances in Efficient Design of Experiments in Economics
Michał Wiktor Krawczyk, John Masson Noble
Excerpt
More Information

1 Introduction and Overview of the Element

Most books about experimental economics start with the observation that experiments can contribute to the understanding of economics, despite some early claims to the contrary. They then proceed with a historical account of the rising popularity and prestige of the discipline; these accounts inevitably conclude that experimental economics is well past its infancy. By now, experimental economics is well established as one of the methods of analysis and therefore we do not think its existence needs to be justified. The goal should rather be to make it as useful a tool as possible.

The key to this is to work on the efficiency of design and inference, so that sample sizes achievable within a reasonable budget can provide rich and precise information. In the current practice of experimental economics, both the design and data analysis plans tend to be rather simple; the researcher may decide on a sample size using a rule of thumb, assign half of the subjects to the control group and the other half to the experimental group, run the sessions, and then conduct some non-parametric tests to see if the dependent variable takes significantly different values in the two groups. Simplicity is a virtue and this approach often works reasonably well.

In other cases, though, more sophisticated methods can yield considerable gains in efficiency. Fields with a longer tradition of conducting experiments have developed a number of techniques that could be applied fruitfully in experimental economics as it matures, but to date this has happened only rarely. One example would be optimal designs seeking to maximise information ex ante (e.g. D-optimal designs). They are employed, for example, in discrete choice experiments conducted by environmental economists (Mariel et al. 2021), but they are not commonly known among mainstream experimental economists.

Then there are various innovations within the profession. Adaptive (aka dynamic) designs (such as the non-Bayesian Perny et al. 2016) and several Bayesian approaches: DOSE (Chapman et al. 2018) ADO (Adaptive Design Optimisation) Cavagnaro et al. (2013) and the approach of Toubia et al. (2013), in which previous responses help to determine what stimuli are likely to be most informative about subjects' preferences, have recently been developed but are not widely adopted yet.

The coverage of these topics in leading texts on experimental economics (e.g. Jacquemet and l'Haridon 2018) and experimetrics (Moffatt 2015) remains patchy. Moreover, promising links between them have been largely overlooked.

In this Element, we explore the pros and cons of the simplistic business-as-usual approach to designing economic experiments. We discuss the pervasive

© in this web service Cambridge University Press & Assessment www.cambridge.org

problem of small, underpowered experiments. We point out the factors that make it easier to run large-scale experiments involving simultaneous manipulation of several variables. These include professionalisation of laboratories and the rise of online experimentation. The latter also makes researchers try to shorten their experiments because, compared with the lab, it is harder to keep online subjects focused for an extended period of time. It is therefore imperative to improve the efficiency of experiments. We highlight developments facilitating in absentia dynamic adjustments of stimuli. We also discuss changes in experimental practice, such as pre-registration of experiments and peer review of designs (so that a paper may be tentatively accepted for a journal before the data is collected), that encourage very detailed planning of the design and data analysis at the onset of the project.

We identify the benefits of the novel approaches in terms of efficient elicitation of preferences of subjects. Wherever possible, we try to quantify them based on the measures developed in the literature on Bayesian optimal experimental design, using simulations, or reporting existing empirical results.

We explore the additional insight that these innovative methods yield into the process by which preferences arise and crystallise. There is abundant evidence that the behaviour of experimental subjects is inherently noisy and context-specific. As a result, stability of findings across methods and trials is often disappointing.

We also discuss the costs and limitations of deviations from the experimental economist’s ‘business as usual’. These include subjects’ potential confusion about more complex experiments and possible incentive compatibility and deception issues in dynamic designs.

We strived to base the Element on a consistent statistical approach. During the process of writing, the second author switched from an agnostic position in the classical versus Bayesian question to a strongly pro-Bayesian position. The Bayesian approach is clearly intellectually more satisfying; as Lindley points out, all statistics has some ad hoc element and the advantage of the Bayesian approach is that this is all encoded in the prior distribution. Once the prior has been established, the whole analysis follows clear mathematical logic. The main problem with the Bayesian approach is (and always has been) the computational complexity, due to the fact that the integrals do not have closed form; Gibbs samplers and Metropolis–Hastings Markov Chain Monte Carlo techniques are required. While the classical approach presents a computational framework that is usually much quicker, it became clear that Bayesian computations are now manageable, even with limited computational resources. Furthermore, it is very rare in experimental economics that one approaches a problem with a genuine complete lack of any prior information; there is

CAMBRIDGE

Cambridge University Press & Assessment
978-1-009-50035-7 — Advances in Efficient Design of Experiments in Economics
Michał Wiktor Krawczyk, John Masson Noble
Excerpt
More Information

usually information from previous experiments on related issues and, when genuine prior information is incorporated, one can reach conclusions, even with a small quantity of new information. This is most evident in the case of dynamic/adaptive designs, in which Bayesian updating is the mathematical notion of choice.

The Element is structured as follows. In Section 2 we discuss some fundamental issues of experimental design and causal inference. In Section 3 we discuss the principles of optimal design. Section 4 deals with choice experiments, while Section 5 describes advances in adaptive designs.

The target group is researchers interested in running economic experiments. We assume the reader is familiar with basic concepts of the method.

2 Causality and Random Assignment

In this section, we briefly discuss the key characteristics of causality and show how cause and effect can be established using various approaches to sampling and random assignment, covering such issues as between-subject versus within-subject designs.

2.1 What Do You Mean by 'Cause'?

Experiments are often, and for good reasons, portrayed as a gold standard to establish causality. But what does it mean exactly that X causes Y? How can we find out that it does? Perhaps Y causes X. Perhaps it goes both ways. Perhaps another variable affects both. Perhaps it is mere coincidence. The ontological and epistemological issues of causality have long been discussed; see, for example, Brady (2011) or Thye (2014) for a user-friendly introduction for social scientists, Mahoney and Acosta (2021) and Woodward (2016) for a more in-depth review of 'causality as regularity' and 'causality as manipulation' types of theories, respectively. Classic books include Spirtes (2001) and Pearl (2009). Here, we report a small part of this discussion which seems to be most relevant for the design of experiments in social sciences.

The pioneer of the investigation of causality, David Hume, famously chose the game of billiards to illustrate it: 'Here is a billiard ball lying on the table, and another ball moving toward it with rapidity. They strike; and the ball which was formerly at rest now acquires a motion. This is as perfect an instance of the relation of cause and effect as any which we know, either by sensation or reflection.' Indeed, the collision, with no contributing or intermediate factors involved, will always cause the second ball to start rolling. Moreover, the physical mechanism is well understood; we also can easily establish that the second ball remains motionless if nothing strikes it. Social sciences tend to

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-009-50035-7 — Advances in Efficient Design of Experiments in Economics
Michał Wiktor Krawczyk, John Masson Noble
Excerpt
More Information

provide us with less-than-perfect instances of cause and effect. Social phenomena tend to be complex and probabilistic and hardly ever have a single cause. In fact, 'everything is related to everything else'. Moreover, the laws proposed to explain social phenomena are often contested and their validity may vary over time and across cultures. Still, the basic principles of defining and identifying causality carry over from the simpler, physical phenomena.

Correlation If A causes B, observing B is more likely if A has been observed, compared to the situation in which A was not observed. Naturally, a similar statement can be made for non-binary variables: if there is a causal link, we expect a correlation – for example, high values of A being systematically accompanied by high values of B. Clearly, this is not a sufficient condition. One of the authors is an avid, if inept, hockey player. The only time he was painfully hit by a puck (actually, twice on the same night!) was just hours after he received one of his Covid-19 vaccine doses. This remarkable correlation did not turn him into a vaccine conspiracy believer. For more examples, try to google for images using 'correlation is not causation' or a similar query to see time series that are very unlikely to be causally linked (say, the number of Ariana Grande's Instagram followers and the number of cases of African swine fever virus) and yet, over a purposefully selected period of time, turn out to be very highly correlated. While we may easily identify such a nonsensical correlation as spurious (and we may remember from econometrics classes that they arise easily between two random walks with a drift), our minds are compelled to see patterns of causal links if they are slightly more plausible.

Worse, correlation is not even *necessary*, in that other factors will often conceal or even reverse the correlation that we would expect, given well-founded claims of a causal link. An example of particular empirical importance is that of the link between price and quantity demanded. As every student of economics knows, demand is (almost always) downward-sloping; consumers will want to buy fewer units of a good if it is more expensive. Dubbed the *law of demand*, it is one of the very few laws of the dismal science that actually hold. But how do we know that it holds? Looking at the correlation between prices and quantities sold is very misleading, even if the demand can always be met (as in markets for digital goods that can be instantly produced and delivered at no cost). The producer, being free to set the price at any level, is expected to react to (anticipated) changes in demand. Thus, the correlation can easily turn out to be positive: the price is relatively high when a large number of consumers are interested in the product.

This confusion between correlation and causation is firmly established in the language. The relation of 'being independent' in a statistical sense is

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-009-50035-7 — Advances in Efficient Design of Experiments in Economics
Michał Wiktor Krawczyk, John Masson Noble
Excerpt
More Information

symmetric: when variable *X* is independent of variable *Y*, then variable *Y* is independent of variable *X*. This is clearly not true in the way these terms are normally used. Most people would agree that the weather on any given day does not depend on the way they dress. The only likely exceptions would be the fans of Murphy's law and *Other Reasons Why Things Go Wrong* who would tend to believe that not taking a raincoat makes the rain more likely. As a side note, many of them could perhaps be less inclined to believe that *taking* a raincoat makes the rain *less* likely, although the two statements are logically equivalent. More importantly, for most people it is obvious that the way they dress depends on the (current and forecasted) weather. Again, from a statistical viewpoint, if there is dependence, it goes both ways. Statistical (in)dependence is thus not an intuitive notion; our minds like to think in terms of *causal* links.

Theoretical Plausibility As in the case of Ariana Grande and swine fevers, correlations in pre-existing data may be purely incidental. With a bit of 'luck' it may even be true of a statistically significant experimental treatment effect, especially when many different tests are run; see Abdi et al. (2007) for the discussion of ways to correct for that. Some theories of causality thus emphasise that a cause must be explainable by a 'law-like statement'. This in itself is difficult to define, but the guiding principle is that a purported causal link is much more convincing if it is predicted by a fairly general theory. If it flies in the face of a theory (and intuition), such as the observation that strangers cooperate more than partners in the public goods game (Andreoni 1988) we need to be wary and conduct careful replications.

Counterfactuals and Manipulability The problem in empirical verification of the law of demand mentioned before was that 'high price' regimes tend to systematically differ from 'low price' regimes in other dimensions affecting quantity sold, so comparing said quantity between the two states does not tell us much about the slope of the demand function. The challenge is to imagine 'the most similar world' and determine what the outcome would have been. The main problem is, of course, that this is typically not directly observable.

Any empirical strategy for identifying the causal link between the price and the quantity demanded using existing data thus requires a component of variation in the prices that can be correlated only with the variation in the price via the law of demand. This is often done with the help of instrumental variables. However, the surest, most direct way is to vary the price randomly, thus by conducting an experiment. In fact, manipulability theories of causality stress just that: X causes Y if exogenous manipulation of X, keeping everything else constant, would tend to affect Y. As Holland (1986) quipped in his highly

CAMBRIDGE

Cambridge University Press & Assessment
978-1-009-50035-7 — Advances in Efficient Design of Experiments in Economics
Michał Wiktor Krawczyk, John Masson Noble
Excerpt
More Information

influential paper, there is ‘no causation without manipulation’. One advantage of this approach is that it breaks the symmetry between the variables, which was characteristic of mere measurement of correlation. Cooling the thermometer will not make you less ill, but treating the illness will lower your body temperature.

While experimentalists may be naturally inclined to endorse the manipulability paradigm, philosophers have raised numerous lines of criticism against it, including anthropocentrism and conflation of ontological and epistemological status of causality. From the viewpoint of a practitioner of social sciences, one major difficulty is that many interesting effects involve purported causes that are not manipulable. Women tend to earn less than men, also when controlling for their education, experience, and other measures of social capital. It would seem natural to say that the mere fact of them being women *causes* their lower wages, although the mechanism involved is likely complex. But exogenous manipulation of biological sex, keeping everything else constant, is unthinkable. In the manipulability paradigm, the sex of an individual can thus hardly be the ‘cause’ of anything. Admittedly, measuring such a causal link is highly problematic even for other approaches. Suppose that the differences in wages disappear if we additionally control for height. Can we say there is no effect of biological sex on wages in this labour market? ‘Just try to be a bit taller’ is not a piece of advice likely to cheer up women unhappy about wage inequalities. Perhaps the relevant ‘most similar world’ here for an average-height man involves being an average-height woman, not a rather tall woman. A discussion of the similarly tricky possibility of considering the category of ‘race’ as a causal variable can be found in Holland (2003).

Timing Another key aspect of causal link concerns timing. We expect the effect to follow (rather than precede) the cause. Sometimes, the effect is significantly delayed, which makes it more difficult to find out that there is indeed a link. For example, it took decades to realise that cigarettes are harmful, because the effect accumulates after years of smoking (although the vested interest of tobacco companies was another important factor delaying the conclusion). At the other extreme, when the reaction is immediate, telling the cause from the effect is sometimes not obvious. For example, many people believe that it is possible to detect that someone is watching us. Titchener (1925), who was probably the first to address this superstition scientifically, suggested the belief could be related to the natural tendency to pay attention to movement. When person A (male) turns around, it is fairly likely that person B (female), initially behind person A, will throw a glance at him. This reaction tends to be non-conscious and immediate (‘system 1’), so it might seem to B that it

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-009-50035-7 — Advances in Efficient Design of Experiments in Economics
Michał Wiktor Krawczyk, John Masson Noble
Excerpt
More Information

was A who turned *in reaction* to her gaze (although in truth, it was herself glancing at A in reaction to his movement). Likewise, A, seeing that B is looking at him (and not knowing this was not the case just a second ago), may infer that he himself has turned *because* he could sense B's gaze. Clearly, in the case a researcher assigns a non-zero prior probability to the supposition that gaze may be detected, controlled experiments with randomised glances must be conducted and indeed *have* been conducted and some even confirmed such an ability, although, perhaps predictably, the methods are contested (see Marks and Colwell 2000).

Another difficulty concerns expectations. The outcome may come before the cause when the *expectation* of the cause precedes it. Much like a rooster's crowing does not cause the sunrise, a war need not be triggered by a market plunge; rather, a war may affect the markets before the first shot is fired because there are good reasons to *expect* it to be fired.

Not surprisingly, timing is also of crucial importance in the design of experiments. Ideally, we want the outcome measure to be elicited immediately following the experimental manipulation. We know that no systematic difference between the experimental group and the control group should occur prior to the manipulation (and if it does, it likely implies there is a problem in the random treatment assignment procedure or that our preparations to implement the treatment are poorly hidden). Exogenous randomisation also rules out the role of expectations. Even if subjects expect *some* manipulation, they usually do not know its nature and, maybe most importantly, they do not know to which group they will be assigned.

2.2 Counterfactuals and Randomisation Bias

Suppose we have p treatments (one of which may be the 'control', i.e. no treatment); the treatments are labelled $1, \ldots, p$, we give treatment j to experimental unit i, and denote the outcome by $Y_i(j)$. In our experiment, we obtain the result of applying treatment j to unit i, but we would like to infer what the outcome would have been if any of the other treatments had been given; in other words, when we observe $Y_i(j)$, we would like to infer the values of $Y_i(k)$ for all $k \neq j$. This is termed *counterfactual*, since for run i we gave treatment j and we did *not* give any of the other treatments.

We can represent the situation by a causal diagram (Figure 1); the outcome Y is influenced by the treatment T and other causes C, which may influence which treatment is given as well as having a direct influence on the outcome. The treatment variable T takes values in $\{1, \ldots, p\}$. We would like to infer that $Y_i(j)$ (the so-called potential outcome for subject i under treatment

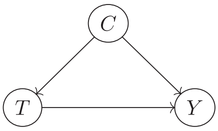

Figure 1 Common cause, treatment, outcome.

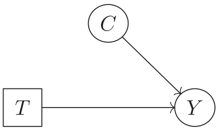

Figure 2 Intervention breaks the link between common cause and treatment.

j – the outcome that would have occurred had subject i been administered treatment j) is independent of T (the treatment actually given), which implies that $\mathbb{E}[Y_i(j)|T = k]$ does not depend on k (and hence that, for each j and k, $\mathbb{E}[Y_i(j)|T = k] = \mathbb{E}[Y_i(j)]$). This is (of course) true if treatments are assigned at random and a *proper* controlled experiment is carried out (i.e. the sample is representative of the population at large and treatments are randomly assigned). This would break the *causal* link between the common causes C and treatment T and result in the causal diagram of Figure 2 (since T is enforced by the experimenter, we denote this by a square node).

Randomisation of treatment means that the choice of treatment is applied irrespective of the values of any hidden covariates that may have a causal effect on both treatment and outcome.

In many situations, though, there are serious obstacles to the construction of a controlled experiment. For example, a sample drawn from college students who *agree* to participate in an experiment may differ in important ways from those who choose not to participate, so the whole sample may be biased. In experimental economics, an important reason for differences is related to the structure of pay-offs. Harrison et al. (2009) hypothesise that the anticipated variance of pay-offs may affect the profile of risk attitudes of the sample of people who sign up to take part. They confirm that announcing a guaranteed show-up fee leads to a relatively risk-averse sample. Therefore, the problem is not whether treatments are assigned in a suitable random manner to a pool of candidates constituting a representative random sample; the *whole sample* of possible candidates may be biased with respect to the prevailing characteristics of the population.

Random assignment may discourage some participants. Levitt and List (2009) point out that in clinical drug trials, it seems much harder to persuade patients to participate in a *randomised controlled* experiment than to persuade

CAMBRIDGE

Cambridge University Press & Assessment
978-1-009-50035-7 — Advances in Efficient Design of Experiments in Economics
Michał Wiktor Krawczyk, John Masson Noble
Excerpt
More Information

them to take a new drug in a non-randomised study. The sample will therefore tend to be skewed for randomised controlled trials.

Framing might matter as well. As pointed out by List (2021), in field experiments, the use of the word ‘experiment’ itself can cause difficulties, while terminology such as ‘trials’ and ‘pilot studies’ may be more acceptable. This applies not only to participants, but also to non-academic partners who are often necessary to run a field experiment. Some of them are prone to respond along the lines of ‘we’ve been in this business for thirty years, we know what to do, and you’re telling us to choose something at random?!’

2.3 Randomisation Procedures

At this point, we introduce another important ingredient, which is how to assign treatments to subjects who have agreed to participate. Many procedures have been proposed for the random assignment of participants to treatment groups. We outline common randomisation techniques, including simple randomisation, block randomisation, stratified randomisation, and covariate adaptive randomisation. We describe the methods, giving advantages and disadvantages.

2.3.1 Simple Randomisation

For this, we simply decide how many subjects are to be allocated to each treatment group; if n_i subjects are to be given treatment i, assign i to n_i labels, put all the labels into a hat, and assign them randomly to the subjects. This technique is easy to implement. The disadvantage is that there may be obvious characteristics (e.g. some of the subjects are male while others are female) that lead to heterogeneity; for relatively small experiments, we may find that there are substantially different proportions of men and women receiving different treatments.

2.3.2 Blocking and Stratified Randomisation

A *block*, as explained by Box, Hunter, and Hunter (2005), is a portion of the experimental material (e.g. two shoes on one boy, two seeds in the same pod) that is expected to be more homogeneous than the aggregate (the shoes of all boys, *all* seeds available, not necessarily from the same pod). By confining comparisons to those within blocks, greater precision is usually obtained, because the differences due to belonging to different blocks are eliminated. Blocking seems to have been introduced by Student (1911). It has similarities to so-called *stratified randomisation*. For analysis of data, they are treated similarly; the difference is that the experimenter assigns subjects to blocks, while the strata refer to covariates, or characteristics possessed by the subjects

© in this web service Cambridge University Press & Assessment www.cambridge.org