Initiating an Experiment

Noushin Nabavi

2020-09-08

Contents

Preface

Welcome to the first edition of initiating an experiment in the public service. I'm so excited to work on this chapter for the experimentation course and work with my vibrant colleagues in the EW2 cohort.

This is a repository to house course materials related to module 2 of Government of Canada's Experimentation Course: *Initiating an Experiment: what do you need to consider before starting an experimental project?*
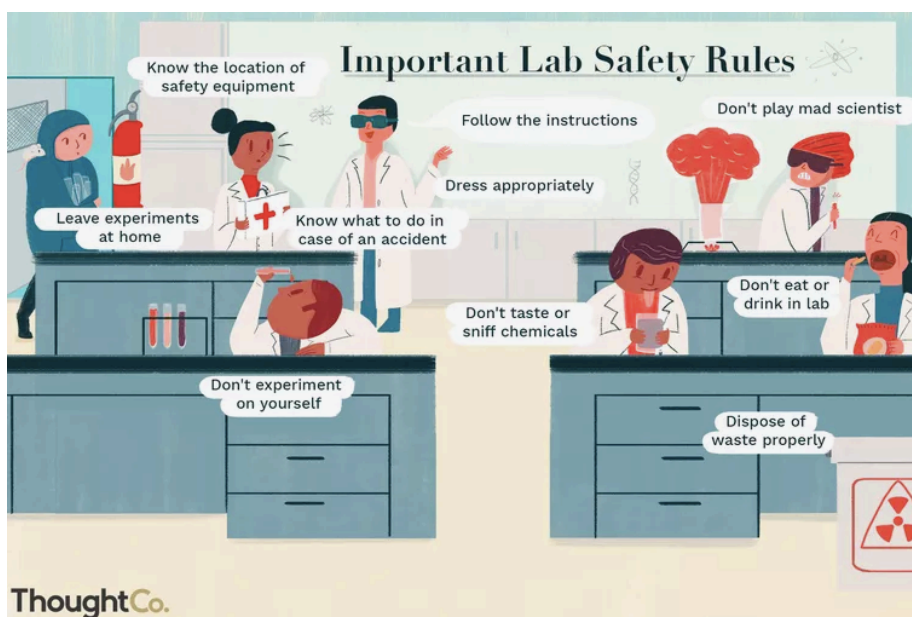
I hope this resource will be useful in clarifying **when** to experiment, **what** experiments are useful and **why**, **which** pitfals to consider, as well as simplifying scientific and technological jargon and common misperceptions. As innovators seeking social good, we have a duty to put our ideas to test and discover what doesn't work, what does, and find out how to improve people's lives.

I am tremendously grateful for the mentorship I have received from the EW2

leading team (Pierre-Olivier Bedard, Dan Monafu, and Sarah Chan) as well as the graphical design work lead by Jordana Globerman.

I will appreciate to know how this course can be improved and what additional topics can be added. I also love to hear how public servants foresee using the experimentation concepts in their projects towards evidence-based transformation of policies, programs, and services.

A big thanks to Preet Chauhan, EW2 expert, for proof-reading and helping with the many examples.

0.1 Prerequisites

This course is intended for anyone with a curiosity and interest in making scientific observations through experimentation, and does not require previous experience of studying the subject.

An interest in experimentation, willingness to experiment, and an apetite for informed consumation of evidence. It means admitting that we don't know all the answers but we can put our ideas to test.

Experimentation also necessitates a more mature attitude from leaders and decision-makers, one that is willing to experiment and learn positively from "good failure" rather than pretending we have all the answers.

A team of knowledgeable partners with expertise in experimentations methodologies with a transparent and "open-by-default" approach.

0.2 Themes

We will explore the following themes in this module to help us answer the question of what we need to consider before starting an experiment: (1) opportunities to experiment in government, and (2) designing experiments as part of existing programs/services.

0.3 Learning objectives

By the end of this module, you will be able to:

- Identify the steps needed before starting an experimental project
- Devise a problem statement for the project and design a research experiment
- Apply experimentation knowledge to specific examples, cases and scenarios

0.4 Outline

Chapter	Title
1	Before initiating an experiment
2	Components of an experimental project
3	Mechanics of endorsement
4	Case studies
5	References

Chapter 1

Before starting an experiment

1.1 Do we need to experiment?

To embark on an `experimentation project`, we really need to consider whether an *experimentation* is needed in the first place. One way of determining this is to consider whether out of the many questions that a team has come up with, are there any that require a specific answer to a very specific question? And is the answer worthy to know? If we are after questions that are not causal in nature, then an experiment is likely not the best fit.

At this point in the **experimentation cycle**, there are more questions than answers and the more questions one asks, the more clarifying answers one will seek on whether an experiment is really needed.

What would experimentation look like in policy or practice? why would an experiment be worth investing in? what data or program performance tracking is already in place? what experimental tools are available and whether and when they might fall short? what are the risks? What could be a risk management strategy?
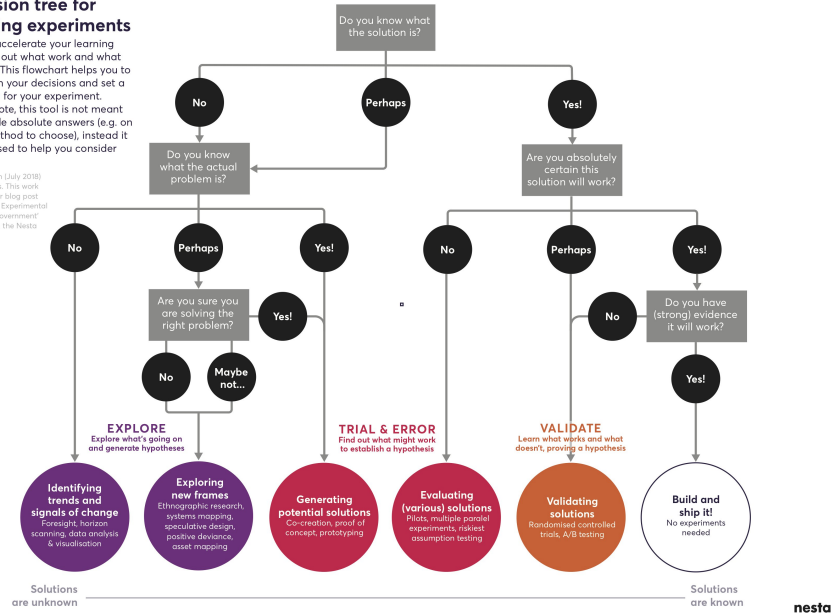
Decision tree

To help make whether an experiment is needed and set one's direction, the following decision tree from Nesta is a great starting point for a project journey. In this journey, only the two red bubbles refer to an experiment. The rest are only explorations or validations. What makes the two red bubbles an experiment? what do they have in common?

Decision tree for running experiments

How to accelerate your learning and find out what work and what doesn't? This flowchart helps you to reflect on your decisions and set a direction for your experiment. Please note, this tool is not meant to provide absolute answers (e.g. on what method to choose), instead it is supposed to help you consider options.

Draft version (July 2018) by Bas Leurs. This work builds on our blog post 'Towards an Experimental Culture in Government' available on the Nesta website.

Do you know what the solution is?

No — Perhaps — Yes!

Do you know what the actual problem is?

Are you absolutely certain this solution will work?

No — Perhaps — Yes! — No — Perhaps — Yes!

Are you sure you are solving the right problem? — Yes!

Do you have (strong) evidence it will work? — No

No — Maybe not... — Yes!

EXPLORE
Explore what's going on and generate hypotheses

TRIAL & ERROR
Find out what might work to establish a hypothesis

VALIDATE
Learn what works and what doesn't, proving a hypothesis

Identifying trends and signals of change
Foresight, horizon scanning, data analysis & visualisation

Exploring new frames
Ethnographic research, systems mapping, speculative design, positive deviance, asset mapping

Generating potential solutions
Co-creation, proof of concept, prototyping

Evaluating (various) solutions
Pilots, multiple parallel experiments, riskiest assumption testing

Validating solutions
Randomised controlled trials, A/B testing

Build and ship it!
No experiments needed

Solutions are unknown ——————————————— Solutions are known

nesta

1.2 The experimentation cycle

> A problem well-stated is half-solved!

To help others understand the problem as we see it, the question needs to be well framed and the goals clearly articulated. To help formulating the `why` as well as the `what`, the following experimentation lifecycle is helpful in that it encourages to reiteratively (1) brainstorm and form vague ideas, (2) group vague ideas, (3) make general observations, (4) hypothesize, (5) determine a model/method to test the hypotheses, (6) do the experiment to test hypothesis, (7) gather data, analyze, and interpret the results, and (8) learn and communicate learnings to stakeholders

The experimentation lifecycle
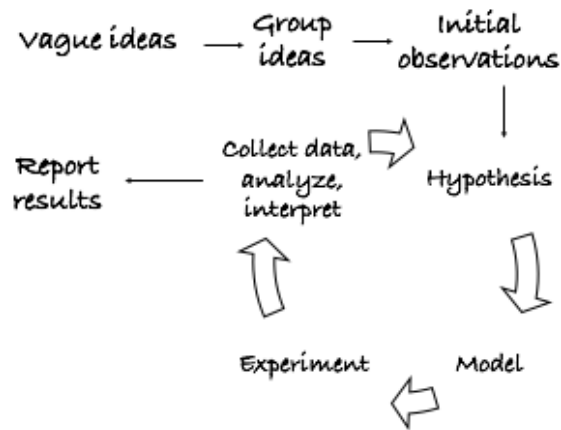
This lifecycle presumes that we can:

- sketch out what information we need to collect (or already have) to get from a vague idea to the hypothesis stage for a planned project
- get invested in the problem before the solution nor in a particular result (any biases will work against us here)
- not get stuck in a fishing expedition (i.e. grouping ideas forever)
- understand the problem well enough to crisply articulate the goals, questions, and hypotheses before building metrics.
- select metrics that will help answer the questions. This can include system parameters, workload parameters, behaviours, etc. `"If all you have is a hammer, everything looks like a nail." - Bernard Baruch`
- identify parameters that affect behaviour or observations and decide which parameters or interactions to study, or vary.

1.3 What can go wrong?

> Don't land your plane in forests, and don't do experimental designs
> before you have considered its drawbacks

Every positive side comes with a negative one, and so no design is really perfect. This is not to discourage us. In a way, George Box's (British statistician) quote that "All models are wrong, but some are useful" applies to experimental designs as well. One single experimental design cannot be comprehensive and all-encompassing. Additionally, we cannot control for the many factors and behaviours that may affect a situation.

Therefore, since all models are wrong, researchers should check the scope of applicability and limitations of their method/model. We should choose the designs that best answer a research question, and not try to tailor the research

question to the method at hand.

For instance, we may decide to base our tests on a set of observations derived from survey findings and not be aware that survey data can badly fail in several aspects: (1) people act differently when they realize they are under study. If asked about questions on sensitive topics (e.g., homosexuality, immigration, abortion, or Donald Trump), people understand there is a "standard" answer and so, they may hide their true feelings and give socially acceptable answers. (2) having a representative sample is difficult and expensive. In social sciences, one of the basis of experiments is recognizing regional diversity (e.g. British Columbia is different from Ontario and Ontario is different from Quebec). This can affect the interpretation of our findings and their generalizability. (3) determining the causal relations between two events is usually the goal of experimentation. However, direction of causal inference in social situations can be problematic. Traditional quantitative methods are particularly ill-designed for causal questions when the answer can go in either directions. A solution here is a carefully designed experiment with control groups
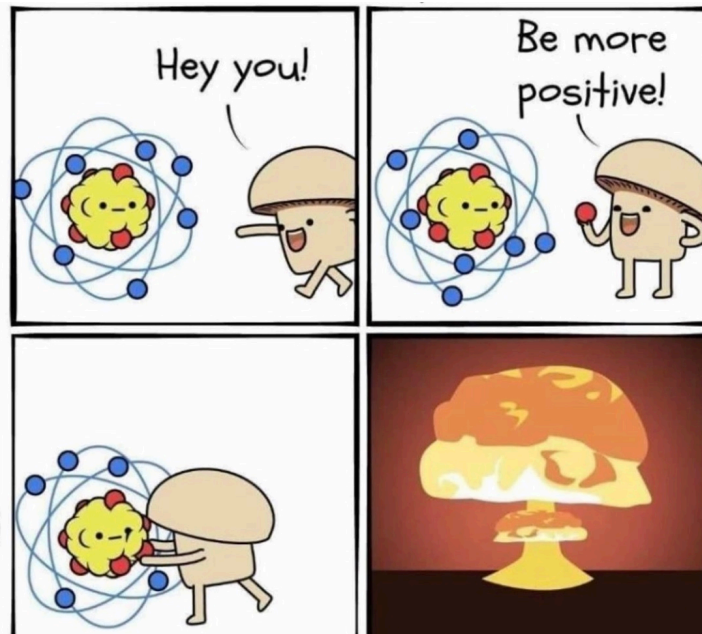
Additionally, statistical models that are used to analyze collected data also work under assumptions. For example, a simple linear regression model requires four assumptions:

```
- E(y) = X
- Independence
- Equal variance ( ²)
- Normality
```

Without checking these assumptions, we are using the wrong models and generating misguided insights or misinformation. Thus, relying on models can also have un-intended consequences.
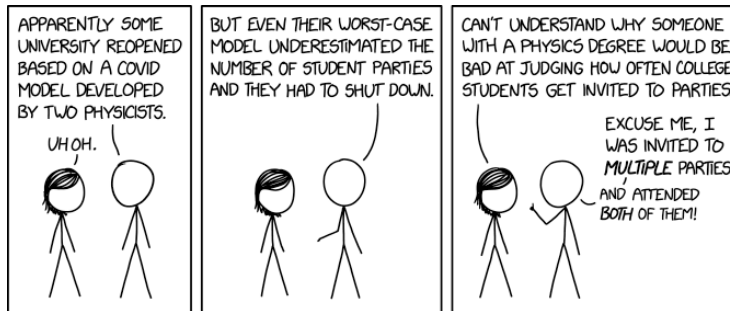
Another instance is not considering degrees of freedom, i.e. measuring how much information we can get from the data: the more variables are included, the less information is left out but our data analysis models can run out of steam if there are too many variables (problem of big data).

To summarize, experiments can be 'dangerous' if we fall into the following pitfalls (not an exhaustive list):

- Devise wrong metrics (i.e. metrics that don't answer the question at hand)
- Have no clear scope (i.e. what are the boundaries for the 'system under test')
- Omit assumptions and limitations of study
- Use unrepresentative metrics, have no comparison groups, or have cross-contamination
- Not recognize the experimental limitations
- Overlook significant parameters that affect the behaviour of a system
- Report average and not variability (fall for tricks of statistics or have no statistics!)
- Have no interpretation of what results mean or overgeneralizing conclusions
- Ignore errors and outliers
- Not consider the ethical issues and scenarios or have informed consent from participants

1.4 It is not all bad

> Evidence-based policy making can be a political ideal

Experimentation is not all bad news. Many breakthroughs and transformations on many fronts from medicine and technology to social changes for good have come about by a willingness to experiment. For example, the following:

There are also many pros to experimentation. For instance, experiments can result in a higher degree of internal validity. Through random assignment of a treatment condition, experimental designs allow us to examine the effect of one variable while keeping other conditions constant. Note that randomization is the key here because it ensures that the treatment and control groups are comparable. Any differences between the two groups can be attributed to the treatment. This is great to know and invest in as a business.

Experimentation can be good even without prior knowledge. This is because sometimes there may not be a theory or theories may fall short to start with. Since an experimentation can directly control how data is generated, the experimental approach can survive and thrive with no previous knowledge. For instance, with no prior knowledge, electoral researchers can carry out a field experiment to examine how different ways of contacting voters would affect the voting turnout.

Experimentation can help clarify mixed results. This is sometimes inherent in observations that are looking at similar phenomena with different measurements or with different data sources. Observational studies often generate mixed results and one way to validate or clarify the existing mixed result is to run an experiment. Again with the example of electoral politics, researchers can chose to run an experiment to detect how campaign spending affects the voters for the incumbent and the challengers differently.

1.5 Communication of intent to experiment

Given that we are considering the option to experiment, we need to have clear and open communications and collaborations with cross-functional teams.

The aim is to get feedback from as many diverse people as possible. These conversations can not only help us decide which ideas to take forward, but also reiterate clear questions based on feedbacks, and eventually implement in an experimentation proposal.

These conversations presumes that there is already an executive level buy-in in place and that stakeholders are invested in the experimentation process. This also assumes that a clear and thoughtful design and implementation plan exists before starting to communicate such that it can be communicated wholly and incrementally with the executives. This is not a tautology and hopefully drive the point that iteration, re-iteration, agility, and an open attitude are key qualities in the initial process.

Executive buy-in usually requires a thorough risk assessment and contingency plans more so in the public service space than in academic environments. Therefore, it is important that the executive is aware of the experimentation cycle which can also be thought of as the problem-solving process and endorses the method, approach, tests, and tools that generate evidence. Some, in fact, argue that experimentation is the creation of something new in the face of uncertainty

and risk.

> It is noteworthy that the word "experimental" has come to mean "innovative" or "radical" rather than simply "untested". The Experimenter's Inventory

Genuine experimentation is about committing to rigorous assessments and evaluation of evidence, not just freewheeling "trying stuff out" or doing things differently and expecting to succeed.

Therefore, even though a methodology is key to answering a specific question, the starting point is having a problem you are trying to resolve, preferably with the social good in mind. The purpose of experimenting is to test key questions and assumptions using quick, low risk experiments.

From the thousands of experiments conducted by Thomas Edison to create the first lightbulb, or the long-running field experiments by Gregor Mendel to examine genetic variability that today underpins modern agriculture concepts, through to trials in medicine, carefully testing ideas in practice is a cornerstone of scientific and technological discovery.

1.6   Experiments in the public sphere

Today, experiments are critical to sectors where innovation and optimization are routine, such as web development, digital transformation, electrical vehicles, etc. This has caught on in business such that the largest financial institutions, retailers and restaurants are also running randomized experiments, along with companies like Google, Facebook, and Amazon running tens of thousands of experiments a year. A/B testing is now the standard means through which Silicon Valley improves its online products. However, in government experimentation remains relatively rare and a new field.

A small but growing movement of policy experimenters are bringing fresh ideas on how to solve public problems. From crafting better services, to making the back-office of government more efficient, new methods and tools need to be used to develop and test policy. In fact, government must rigorously and systematically put policy to the test – or risk stagnation.

`In the context of government agencies, experiments aim to evaluate a program, policy, or service and test an idea or innovation by investigating what difference it has made or will make for the people it is aiming to help.`

Like laboratory experiments, public sphere experiments also need a control group to test an innovation against "business-as-usual". This doesn't have to be a large trial like testing a drug and can be fast and flexible. In fact, the best experiments start small and as a prototype before they are extended.

For example, the World Bank advocates for "nimble randomized control trials". They funded nimble evaluations on how best to improve the take up of

health insurance in Azerbaijan, expand the use of contraceptives in Burundi, and support teachers to deliver tailored education to children affected by war and displacement in Lebanon.

1.7 Summary: deciding to experiment

- Do you need to experiment? Why or why not?
- Find a behaviour, program, policy, or service to test
- Try out of the box thinking to brainstorm, make observations
- Look for natural experiments
- Talk to experts and get feedback
- Think small and short term
- Start with a proof-of-concept question and hypothesis
- Keep it simple and try to test one thing at a time
- Measure everything that matters
- Have control and treatment groups when possible

Chapter 2

Experiment components

2.1 The problem statement

> Experimentation as a problem-solving process carried out under controlled conditions in order to discover an unknown effect or law, to test or establish a hypothesis, or to illustrate a known law

Experimentation is described as many things: a method, an approach, a test, or a tool to generate evidence. All of this is true, but experimentation is first and foremost a problem-solving process.

The starting point for your experiment should not be the methodology, nor a predetermined answer; it should be the problem you are trying to resolve.

WHAT EXACTLY ARE WE TESTING.....?

To help us chose the problem statement, we can go through the following steps:

Step 1: Choose a topic 1. What do I find interesting about the subject? 2. What is known about the subject? 3. What is missing and what are the gaps?

In theory, the topic could be in any field (for instance, agriculture, but also

fiscal policy) and at any level (for instance, it could affect 20 people, or it could reach most of Canada's population). It could touch upon the realm of policy, program, service delivery, regulatory and internal services.

The problem doesn't always need to be causal in nature, in which case perhaps an observational study is better.

Step 2: Narrow the topic 1. What do you need to know more about on the topic? 2. Are you interested in social, political, economic, gender, religious issues related to your topic? - Find a "slant" on your topic and determine the root causes (physical, social, or organizational causes) - Develop possible solutions and select a solution 3. Will the results reveal something new or unexpected? 4. What is in scope and what is out of scope? 5. Can you clearly define hypotheses (If…then…) and explicitly state research questions?

The problem statement you develop should be clear and precise, but also scoped into a relevant policy area that is within your reach (in terms of jurisdiction, for instance). A problem that has too little impact to be deemed important is likely not a great fit, nor is a problem with little decision-making implications very useful to be explored. Similarly, problem statements that are too vague need to be modified so that they're scoped appropriately.

Step 3: Find Resources 1. How do we design a research proposal that builds on existing knowledge to address critical questions? 2. The key here is getting to know what is already done on the topic and how can our design improve or build on current knowledge. 3. Perform a systematic literature search by using the keywords you have compiled and use them to search for books in Library Catalogs, blogs, or articles in online databases 4. Consult with experts and seek feedback

In a government context, the right scale and scope of the experiment will be dictated by what decision-makers need to know coupled with the various practical constraints that inevitably come along through the design process.

In general, this step will help us make the proposal more mutually exclusive and collectively exhaustive, **MECE**. By the end of this process, more so that before, we will be able to make statements that do not overlap in content and fully describe the problem at the highest level.

Step 4: Seek collaboration 1. Make sure the question is one that other people can get behind and support 2. Establish collaboration agreements and executive buy-in 3. Peer-review for clarity, scientific accuracy, and feasibility 4. Does the team have the expertise required to complete the project? If not, who else needs to be on team

Figuring out the problem will take time, and needs to be done in consultation with as many parties directly or indirectly potentially impacted by the experiment's results. Before starting the problem definition stage, you should take time to discuss alongside colleagues, management and others (experts and practitioners, current users) what exactly you wish you fix.

Each of the steps above should be completed before moving on to the next one. However, steps can be repeated. For example, if you're on the third step, you can still return to the previous step, and redefine the problem.

2.2   Determine interventions

It's important to be clear what the different experimental designs can and can't tell us. What sets observational studies apart from experimental studies is that there is a component of **intervention** in experimentation.
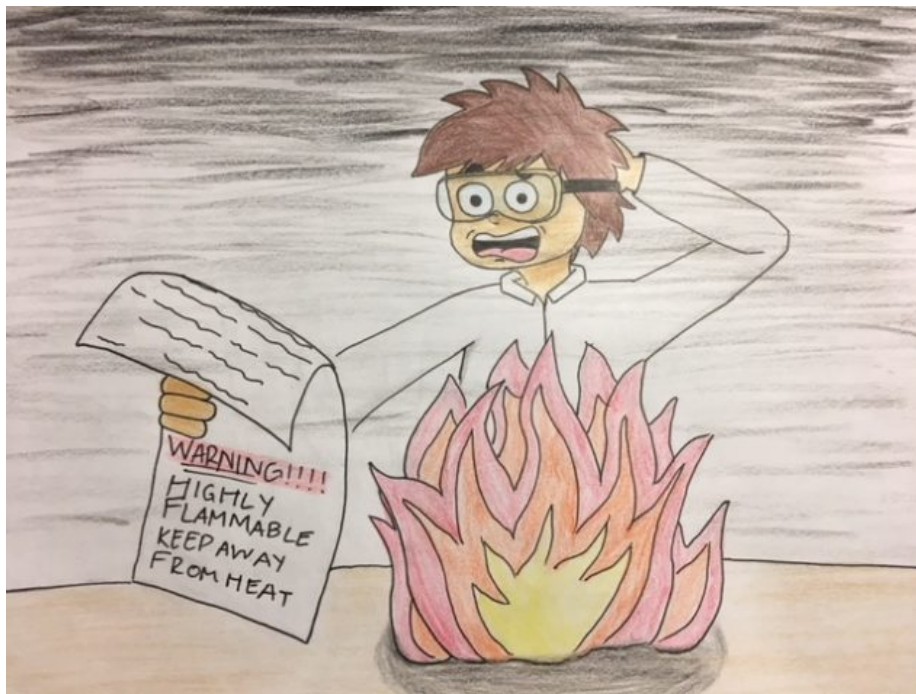

Only certain experimental designs are helpful for learning about impact and effectiveness of an intervention. Therefore, we need to select a design that best tells us whether or not our new idea is really making a difference. This approximates a controlled experiment of basic science. When setting up an experimental design, we need to think about what the intervention (treatment) might look like in our program/ operation. It is helpful to think of interventions as steps that help us evaluate the direct impacts of treatment or preventive measures on a situation.

When thinking about interventions, it is helpful to consider the two following points:

Deciding what to measure: using a government contractor example, consider what kind of data you'd need to answer your key question. In this case,

you'd need to know the number and cost of current staff and the percentage of time they spend on necessary business functions. In answering this question, you likely need to answer many sub-questions (e.g., Are staff currently under-utilized? If so, what process improvements would help?). Finally, in your decision on what to measure, be sure to include any reasonable objections any stakeholders might have (e.g., If staff are reduced, how would the company respond to surges in demand?).

Deciding how to measure it: thinking about how you measure your data is just as important, especially before the data collection phase, because your measuring process either backs up or discredits your analysis later on. Key questions to ask for this step include: - What is your time frame? (e.g., annual versus quarterly costs) - What is your unit of measure? (e.g., USD versus Euro) - What factors should be included? (e.g., just annual salary versus annual salary plus cost of staff benefits)
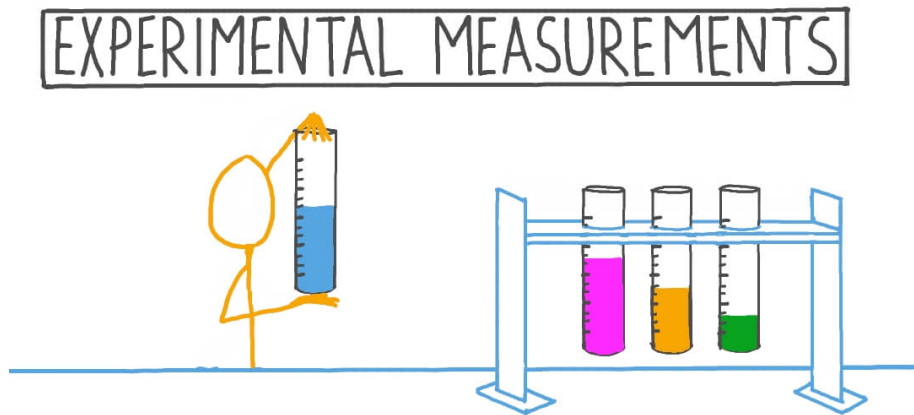

2.3 Determine outcomes

When designing the intervention, it is also important to determine what you want to measure and what the hypothesized effects might be. This will illuminate what data we need to collect, set targets, and develop counter-measures. As part of this, we would want to identify the techniques we will need to perform the measurements and use the collected data for analysis.

An outcome measure, endpoint, effect measure or measure of effect is a measure which is used to assess the effect, both positive and negative, of an intervention or treatment. Measures can often be quantified using effect sizes. They can also be thought of as providing a score, an interpretation of results or at times a risk categorization of study groups. Prior to providing any intervention, an outcome

measure provides baseline data on variable.

Also of note is that experimental studies are usually randomized, meaning the subjects are grouped by chance and we study what happens to people in each group (treatment and control). Any difference in outcomes can then be linked to the intervention.
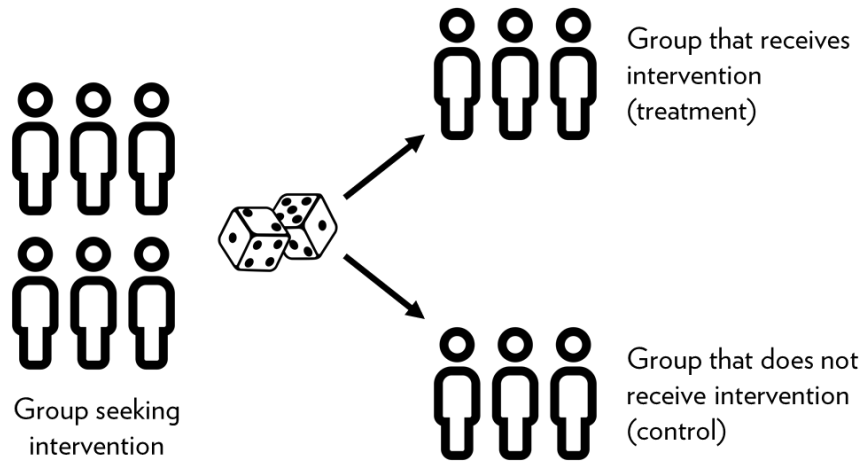
For instance, a well designed randomized control trial (RCT) provides a strong evidence that a given intervention can postulate effectiveness or safety. A RCT is best used when we need to draw conclusions on causality but there are many other experimental designs we can draw on and this will be discussed in Module 3.

> It helps to think of outcome measures through visualization of outcomes / prototypes

As well as increasing our understanding of what experimental design to use when, there is scope for experimenters to get much better at using different tools in combination to innovate more effectively. Prototypers, for instance, could use low-cost randomized trials like A/B tests or nimble trials to evaluate prototyped products or services.

Prototyping emphasizes front-loading risk and creating a solution with a better chance of success through stakeholder engagement. Using plastic Lego bricks to build prototypes of engineering products is a low-fi, low-resource way of making early operational or design issues obvious. But it won't tell you whether or not the new system works in real life, or at scale.

When designing prototypes, we should think of implementing an intervention to measure the outcomes. For this, we need an independent variable or the factor that is manipulated, changed, or intervened. This is usually a variable that is placed on x-axis when grouping. We also need a dependent variable or

the factor that is being measured and usually placed on y-axis during grouping and visualization.

2.4 Experimental design(s)

Next is to determine if randomization is possible, what the sampling unit and approach might be, how the control group might be set up as well as how many observations are needed based on power calculations.

Randomized experiments, or quasi-experimental designs, sometimes couched in deadly technical language, are uniquely valuable. Despite their unfashionable status among some policy wonks and evaluators, we will also highlight the value and utility of Quasi-Experimental designs in the next module. Other approaches may be also be suited to finding out different things, at different stages of developing a policy solution.

Regardless of what type of experiment we chose to do, it is helpful to think about the following variables:

- A control group is a group of 'test' items in an experiment. The control group will be used to compare with the experimental group
- The control group doesn't get the treatment
- An experimental group is the group(s) of test items where only one change (called the experimental or independent variable) has been made
- The experimental group gets the treatment
- The experimental group may have dependent or independent variables

In the case when a real control group is not available, a synthetic control group or a more inventive statistical methods may be useful.

It is also in this step that we determine what information we want to collect, set a timeframe for data collection, and determine the data collection methods.

Source of error considerations - Mind the constants: the conditions that are kept the same for control and experimental groups - Not controlling for factors or parameters that are kept the same in both control and experimental groups can result in error - Type I error ( ) is the rejection of a true null hypothesis (*"false positive"* finding or conclusion, e.g. "an innocent person is convicted") or finding a difference when a difference does not exist - Type II error ( ) is the non-rejection of a false null hypothesis (*"false negative"* finding or conclusion, e.g. a guilty person isn't convicted) or not detecting a difference when one actually exists

		Reality	
		No effect	Effect
Study	No effect	✓ Fail to reject H_0 when it is true	Type II error (β) Fail to reject H_0 when it is false
	Effect	Type I error (α) Reject H_0 when it is true	✓ (Power = $1 - \beta$) Reject H_0 when it is false

Sample size considerations Sample size measures the number of individual samples measured or observations used in a survey or experiment. For example, if you test 100 people for COVID, your sample size is 100.

In general, we want to: - Maximize sample size (n): the larger the number of test items the more accurate the estimate - Use representative groups: the samples must reflect the natural variation in the population. - Use random or systematic sampling to reduce inherent bias in data.

For sample size calculations, we need a hypothesis test, a significance level for the test, the smallest effect size ( ) that is of scientific interest, and the intended power (power = $1 - $) of the test.

There are often softwares that help statisticians decide what sample size and power to chose for a set of sample observations. In general: - The price of increased power is that as standard deviation goes up, so does the probability of a Type I error because - As sample size increases, so does the power of the significance test. This is because a larger sample size narrows the distribution of the test statistic

Statistical power is a measure of the likelihood that a researcher will find statistical significance in a sample if the effect exists in the full population. Power is a function of three primary factors: sample size, effect size, significance level.

2.5   collect and analyze data

At this point, we have a problem, and intervention, a research plan, and data collection strategies. Therefore, we are ready to start implementing the experimental design to collect the data and then analyze it.

We can treat data as qualitative (descriptive) or quantitative (quantitative values/ numbers). Before you collect new data, determine what information could be collected from existing databases or sources on hand. Collect this data first. Also determine a file storing and naming system ahead of time to help all tasked team members collaborate. This process saves time and prevents team mem-

bers from collecting the same information twice. If you need to gather data via observation or interviews, then develop an interview template ahead of time to ensure consistency and save time. Keep your collected data organized in a log with collection dates and add any source notes as you go (including any data normalization performed). This practice validates your conclusions down the road.

Quantitative data comes in the form of numbers, quantities and values and describes things in concrete and easily measurable terms. Examples include the number of students who applied to a grant, the rating a customer gave a product out of five stars and the number of times a visitor spent on the website.

Qualitative data is descriptive, rather than numeric. It is less concrete and less easily measurable than quantitative data. This data may contain descriptive phrases and opinions. Examples include an open ended review a customer writes about a product, an answer to an survey question about what type of applications a customer likes to fill and the conversation an applicant has with a service representative.

Regardless of what kind of data we collect, a proactive data management strategy will make storage, retrieval, and analysis easier.

The analysis phase is also as crucial because it turns raw data into valuable insights that we can use to enhance strategies, products and business decisions. We can often couple data analytics tools to raw data (e.g. power BI softwares, etc) to help with analysis. Visio, Nvivo, Minitab, Stata, and Microsoft Excel are other great tools and software packages for advanced statistical data analysis. The goal of analysis of the data to really to test the effects of the intervention we wanted to measure (e.g. comparing the differences between groups who got a treatment and those who didn't).

Devote some time to review the results. What happened after you implemented the interventions? What worked, what didn't, and what did your solution improve? Analyze if your actions made the required impact and if you addressed the root causes of the issue. It's also time to look for improvements in the solution and to plan ongoing monitoring. You can also analyze what you've learned and what still needs to be learned when it comes to problem-solving processes and skills.

As you interpret the results, ask yourself these key questions:

- Does the data answer your original question? How?
- Does the data help you defend against any objections? How?
- Are there any limitation on your conclusions, any angles you haven't considered?

Once you have uncovered the patterns and insights, you can start reporting back the findings to stakeholders. This step is also reiterative because as you report back, you can adapt the information. Through the presentation of the results, you disseminate the evidence that identities key learnings and adapt the program/ operation based on what the evidence dictates.

The goal of sharing the results and reporting back the process to executives is to reach consensus, generate additional learnings, and develop a better implementation plan.

Problem-solving tips

Problem-solving is a process of constant improvement and reiteration. Don't expect the perfect solution from the start or that the problem won't appear in the future. We should make efforts not to avoid problems because they're part of the learning process. If you adopt an attitude in which you focus on finding solutions every time new challenges emerge, you'll save yourself a lot of time and stress.
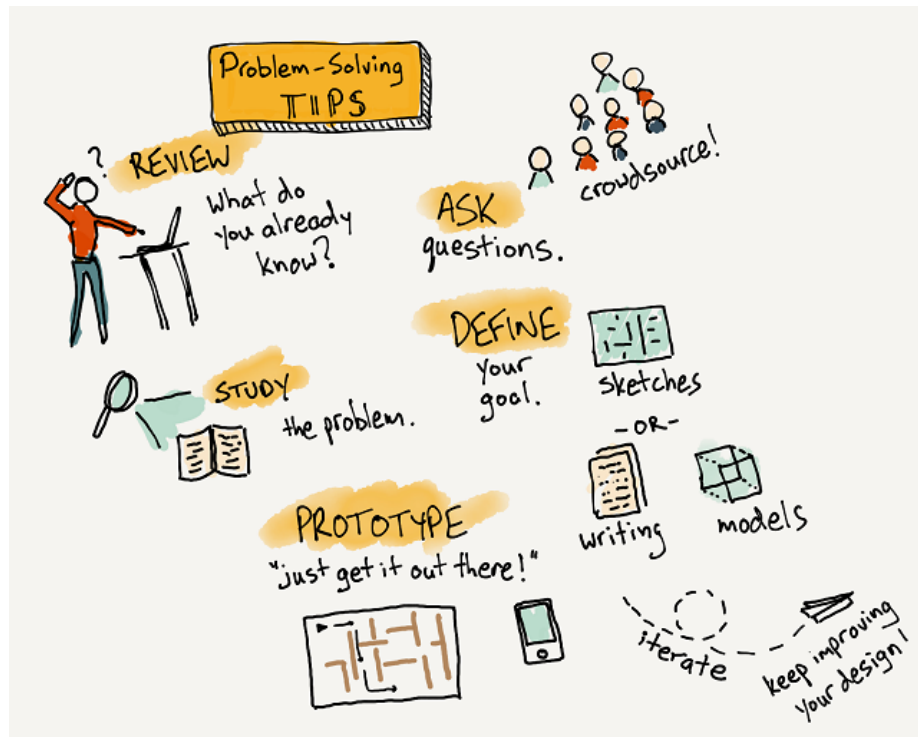
Chapter 3

Mechanics of endorsement

3.1 Knowledge co-production

Communicating the project with intention to gain the trust and cooperation of project team is an imperative step prior to starting an experimentation process. More and more, we are moving towards knowledge co-production models as one person isn't capable of delivering complex multi-faceted projects. To achieve co-production, clear and intentfull communications with team members can affect the success of the project. Buy-in can be difficult to quantify, especially within various stakeholder categories. However, without full buy-in from team members, sponsors and other stakeholders, projects may seem to be progressing smoothly and then suddenly take a sharp turn for the worst, risking the final deliverables and team satisfactions.

Below are the main steps for sustainable co-production research.
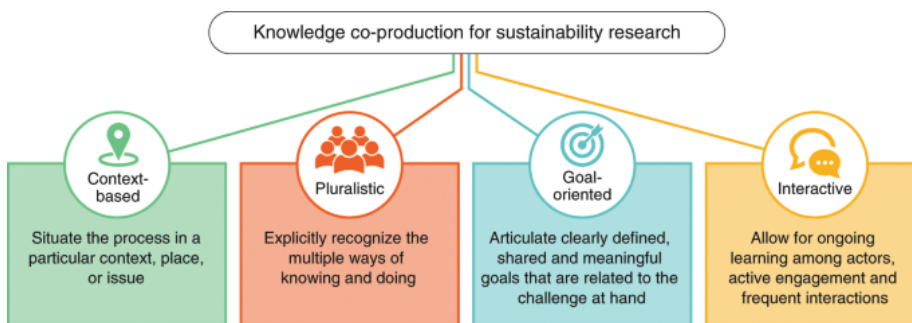
To improve communication with team members at each stage of a project and to make sure they feel like an important part of a team, the following points may be helpful:

1. Identify what motivates the project team and each member individually

2. Focus on telling the truth about the project, even when it isn't what team members want to hear
3. Make sure all team members understand their contribution to the project
4. Reaffirm goals and communicate progress throughout execution
5. Remain consistent
6. Provide positive feedback throughout the process and after the project ends

3.2   Executive buy-in

Once an experimentation plan has been devised, you will need to make a good case as to why the strategy deserves a dedicated budget and resources and what sets this apart from other program initiatives already in place.
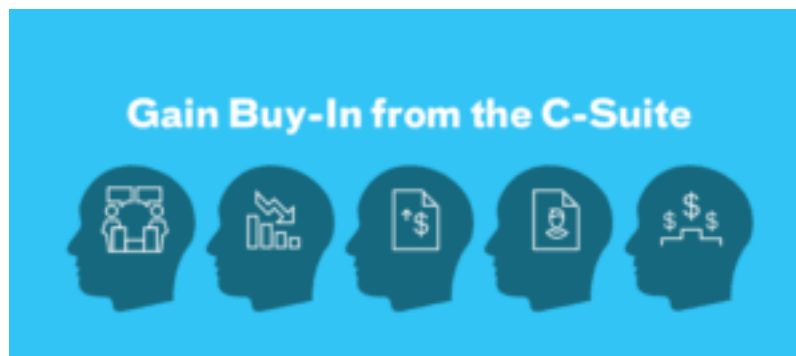
So how do you do it? Below are 3 steps to help in making a convincing case to obtain executive buy-in for your proposal.

1. Figure out who you should be talking to, that is, identify the relevant stakeholders in your organization who have the authority to sign off on a new strategy, allocate budget resources and designate personnel. You can make a list of the people you want to target. Next, you can ask what business needs do they face and what problems does the organization have that an experimentation can solve? Think like a marketer present your plan as a solution to those specific problems. It is also helpful to find a personal advocate within the organization to function as a change agent—someone who believes in you and has enough clout to advocate on your behalf. This will help give your experimental plan some authority when you present it to the executives.

2. Show your accomplishments and what the experimentation can do. This precludes getting to know your organization's analytics capacity and correlating your experimentation success with the real results already existing. For instance, you can communicate that "you have been doing X, and it's been contributing to Y". Evidence-based plans are a powerful formula for getting executive buy-in. What if your community is relatively new, or

not yet in existence? If this the case, you can benefit from the experience of others and researching a cache of other success stories ready to be applied to your programs. The question the executives will want to know is, how will you make another's success your own success. This requires an experimentation plan.

3. Lay out the experimentation's strategic plan for success. Once you have identified key stakeholders and worked on a compelling vision of how you can solve a key problem, you can start sharing that strategic plan. Your plan should, at a minimum, answer the following questions:

- What specific actions must you take to build a community for your research plan?
- What resources do you need to take those actions?
- What results—concrete and abstract—do you hope to achieve?
- What are the risks? how are the risks managed?
- How will you measure your results against your actions?
- How will you quantify the value of this process?
- Are your expectations realistic? Prove it.
- What are your performance deadlines?

3.3 Communication matrix tips

Communication of the intent to experiment and the experimentation processes is like an onion. We can start in the inner layer with peers and colleagues and close supporters who can understand the message. From there, your secondary networks help build momentum. Once it's clear your science resonates with people, then you can reach a larger audience such as the executive or to press and media.

To relay the necessary information about the experimentation plan to stakeholders/executive and to reach consensus, the following points can be considered:

- Why do we need to answer this scientific question now? (*importance*)
- Has this question been answered? Has it been attempted? (*novelty*)
- What's the risk, and what's the potential upside? (*impact*)
- Is the design of the experiment sound? how? (*design*)
- What are the timelines and milestones? (*logistics*)
- What makes this researcher/research team uniquely qualified? (*qualifications*)

3.4 Code of ethics

Concerns about misinformation and the inability to distinguish the contrived from the authentic is growing be it in journalism, marketing, finance, or medicine.

Unlike traditional *observational* methods where one makes observations about the world, the *experimental* approach learns about the world by measuring how it reacts to interventions. Experimental studies raise a host of ethical concerns that social scientists are not always well equipped to deal with. The absence of clear principles to handle complex issues and address ethical implications and unintended consequences can leave teams in a difficult situation, above and beyond whether they receive formal research approval.

There are, however, opportunities to experiment even on difficult social issues such as interventions in poverty policies, drug use, or healthy-living interventions. Creating a control group can carry its own risk. If, for example, we are experimenting with a new benefits scheme, having a control group allows us to test whether the scheme has actually achieved its goals, but it also means depriving a group of citizens of an improved service, at least initially.

Broadly, ethical concerns can fall into the problem of **audience** (what are the issues), **agency** (who is responsible for interventions), and **consent** (are participants informed and do they consent). Some other points to consider include the following:
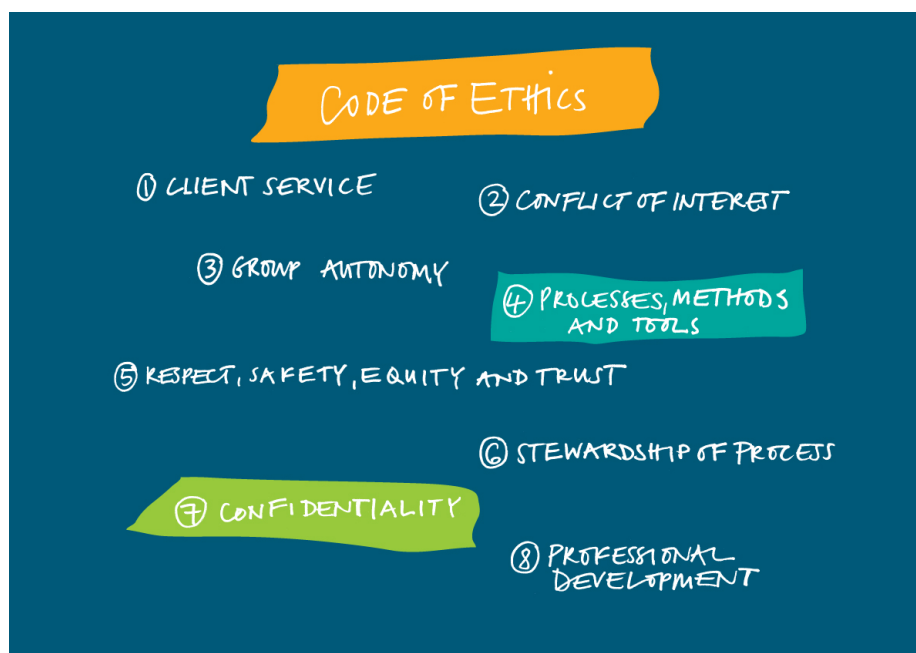
To get around this, a phased policy roll-out can create what's called a "waiting list" experiment where the control group are the soon-to-have-innovation group. Everybody eventually receives the new policy innovation, negating the risk of push-back from the public. This approach was used by the UK's Ministry of Housing, Communities and Local Government in their first ever randomized trial on community integration. In 2016, they tested whether English language

training could help immigrants engage more in their community.

The results were impressive. They found significant improvements on many social integration outcomes, such as new friendships formed with people from other cultures, and attending more health appointments. The trial helped the Department put together new plans for their 2018 Integrated Communities including a network of conversation clubs and a new English language fund.

Ethical considerations: - Inform users and the study population and get consent of participation - Establish information sharing agreements and memorandums of understanding - Identify situations where experimentations are appropriate and relevant - Perform privacy impact assessments/agreements - Get ethics committee approvals - Check Institutional Review Board (IRB) approval requirements prior to the launch of the experiment

Chapter 4

Case study

For illustration purpose, we are going to use a contentious example. Following the 2016 U.S. presidential election, many have expressed concerns about the effects of false stories, *"fake news,"* as it has been dubbed, circulated largely through social media on the election results.

Many have speculated that the exposure to fake news has shaped people's political inclinations in the presidential election. This group argue that the falsified information delivered in a such dramatic way shifts political inclinations. Others, however, argue that consumption of information is more of a selection process: voters choose their sources of political information on the basis on their existing political preferences. As a result, fake news only reinforces the voters' results but will never change their beliefs.

4.1 Background research

Recent evidence shows that:

(1) 62% of US adults get news on social media (Gottfried and Shearer 2016)
(2) the most popular fake news stories were more widely shared on Facebook than the most popular mainstream news stories (Silverman 2016)
(3) many people who see fake news stories report that they believe them (Silverman and Singer-Vine 2016)
(4) the most discussed fake news stories tended to favor Donald Trump over Hillary Clinton (Silverman 2016)

Putting the above observations together, a number of commentators have suggested that Donald Trump would not have been elected president were it not for the influence of fake news (for examples, see Parkinson 2016; Read 2016; Dewey 2016).

4.2   The problem statement

Did fake news affect the election result by making it more difficult for voters to infer which electoral candidate they prefer?

Hypothesis: Fake news is independent of election outcome

4.3   Scenarios

In this example, there are three potential causal relationships.

1. Fake news affects voting: Fake news carry certain pieces of information in a way that changes audiences' preferences.

2.   Political affiliations determine consumption of news: People with certain preferences and orientations choose to watch fake news. So, fake news only reinforces but won't change the existing beliefs.
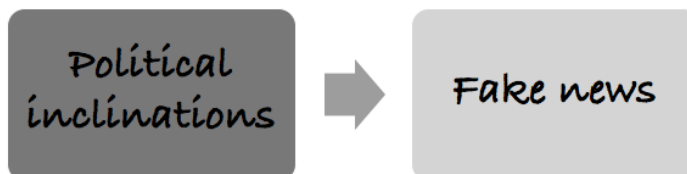
3. Two-way causation: Political preferences determine which pieces of political information to receive. In turn, it further shapes the existing preferences and changes voters' behaviours. To test these scenario, we can apply a simple experimental design and test it.
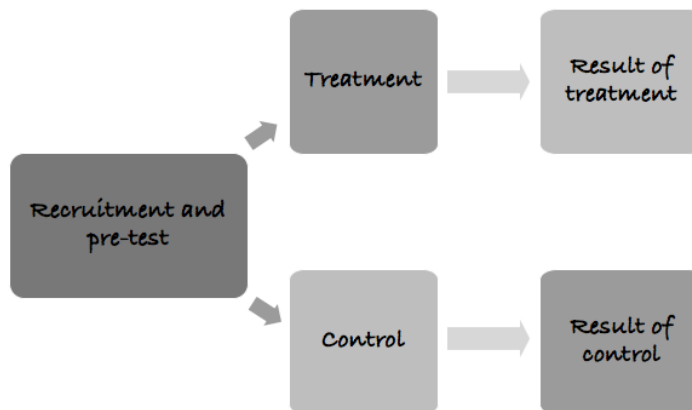
4.4 Experimental solution

Stage one: Recruit and randomly assign participants into treatment and control groups; pre-test.

Stage two: Expose the treatment group to fake news; apply placebo to the control group.

Stage three: Record the results of the treatment and control groups. Due to the randomization process, we can control for any previous differences between the treatment and control groups. Therefore, any differences between Result 1and Result 2 are attributable to the presence of fake news.

Chapter 5

References

- City of Vancouver Solution Lab's Principles of Experimentation, adapted from Nesta's Innovation Playbook
- Nesta's Competency Framework for Experimental Problem Solving
- States of Change's Core Elements of Innovation
- Tatyana Mamut's eight Innovation Elements
- The Moment's Culture Scan
- Innovation Designer Capability Map
- Alliance for useful evidence
- Better public services through experimental government
- The What Works Trial Advice Panel
- Integrated Communities Strategy green paper
- The impact of Sure Start Local Programmes on seven year olds and their families
- National Evaluation of the Troubled Families Programme 2015 - 2020
- Social Media and Fake News in the 2016 Election
- Reflections on the Ethics of Social Experimentation