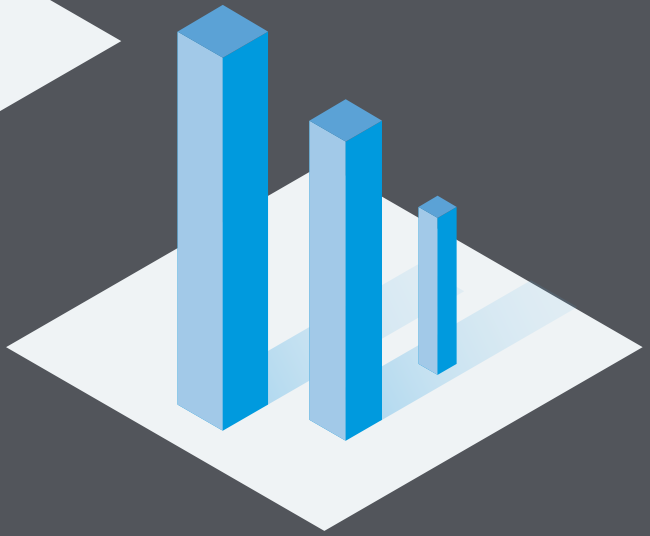
Explore

Four simple ways to map and unpack behaviour

Rory Gallagher, Alex Gyani, Cameron Tan & Karen Tindall



THE
BEHAVIOURAL
INSIGHTS
TEAM

A bit about us

At the Behavioural Insights Team (BIT), we use behavioural insights – findings about human behaviour from the inter-related fields of behavioural economics, experimental psychology and social anthropology – to help organisations around the world achieve social purpose goals. We have spent the last decade partnering with organisations to help address an array of issues as diverse as economic growth, public health, energy and sustainability.

If you would like to learn more, contact us at: info@bi.team

Why has the Behavioural Insights Team written this guide?

Over the course of the last 10 years, we have been asked for a guide on how to do Explore work by a number of our partners. In some cases, these partners have been policymakers who are familiar with qualitative research, but want to know more about how they can use their administrative data to understand their policies. In other cases, our partners have been foundations and not-for-profit organisations who are keen to learn more about their beneficiaries. Or sometimes our partners have been private sector organisations who are keen to see how they can increase their social impact by understanding the context in which they are operating. We have used the term 'policymaker' to describe anyone who fits into this category. Being a policymaker does not mean that you are in government.

While our definition of a policymaker is broad, we have written this guide with a specific audience in mind: someone who wants to develop better policies, services and products, and acknowledges that it is important to Explore the context in which you are operating before trying to change it. If you are still on the fence about this point, we hope that we can convince you otherwise.

Once we have convinced you of the importance of Explore work, we will give you a whistle-stop tour of how to use the fundamental Explore tools. There are many excellent textbooks on how to conduct effective interviews and focus groups, how to analyse administrative data, how to write surveys and how to undertake participant observations.[1] We have cited many in this report and encourage you to read them. However, we felt the need to write this document for two main reasons:

1. We think that policy making can be vastly improved by good Explore work
2. We think that the behavioural insights approach can offer a new perspective on Explore work

Policies, services and products can be improved by good Explore work

The behavioural insights (BI) approach uses evidence of the conscious and nonconscious drivers of human behaviour to address practical issues.[2] Since 2010, we have focused on policy problems. These have largely been in the public sector, but the behavioural insights approach is increasingly being used in the private and third sector.

Historically, whenever BIT has described its projects, these descriptions are usually accompanied by an overview of a Randomised Controlled Trial (RCT)[3] testing whether a behaviourally informed solution has had an impact on a particular policy issue.[4] From these descriptions, a reader might assume that the idea for the solution leapt from the page of an academic journal into reality or that a behavioural insights practitioner used some form of divination to see that a specific finding from the literature would solve the problem at hand.

This assumption does disservice to the many hours of work that goes into understanding the specific policy problem and its context. Effective interventions are created by a process of matching a theoretically driven understanding of a problem with a theoretically driven intervention.[5] This view is also widespread in the development of strategy in the private sector.[6] The inclusion of behavioural economics in the policymaking canon has done much to ensure that the model of human behaviour used by policymakers is more accurate. But BI practitioners need to ground this knowledge in the day-to-day realities of the specific context that they are trying to change.

We have helped thousands of behavioural insights practitioners develop case formulations for their policy problems. Through this work, we have identified some of the fundamentals of doing good Explore work. This paper aims to help other practitioners apply them to their own policy problems.

The behavioural insights approach can offer a new perspective on Explore work

While effective Explore work is good for policymaking in general, using the behavioural insights approach can make it even more effective. The impact of the contextual environment on behaviour has long been noted in the behavioural sciences. In 1936, the German-American psychologist Kurt Lewin argued that someone's behaviour is a product of the interaction between their situation and their personality.[7] Lewin defined personality broadly, comprising a person's past, present, future, motivations and desires.

Since Lewin, more work has been done to understand the specific ways in which the interaction between the person and the situation affects people's behaviour.[8] The lesson from this work is that, if you want to understand people's behaviour, you need to understand: 1) their thoughts, attitudes, values, personality traits, intentions and motivations; 2) the social and physical environment in which they operate; and 3) how the two interact with each other. To make things more complicated, the interaction between the person and the situation is often filtered through a set of cognitive biases.[9]

A key part of a behavioural approach involves identifying these contextual and environmental factors and using them as the basis for solutions. The exploratory tools that we describe in this guide allow you to peek into people's thoughts, attitudes, values, personality traits, intentions and motivations. Interviews, focus groups and surveys are especially good at this. Other methods are better at helping you understand people's behaviour in the context of their environment. For example, observations and participation and administrative data analysis.

These tools are best deployed with some knowledge of the behavioural biases that affect us as researchers and as human beings. We often state that we use a behavioural lens to understand behaviour. By observing people's behaviour through this lens, we might be able to develop better policies, services and products.[10]
Some behavioural biases or behavioural concepts that might help us understand the interaction between the person and the situation are shown in the table below.

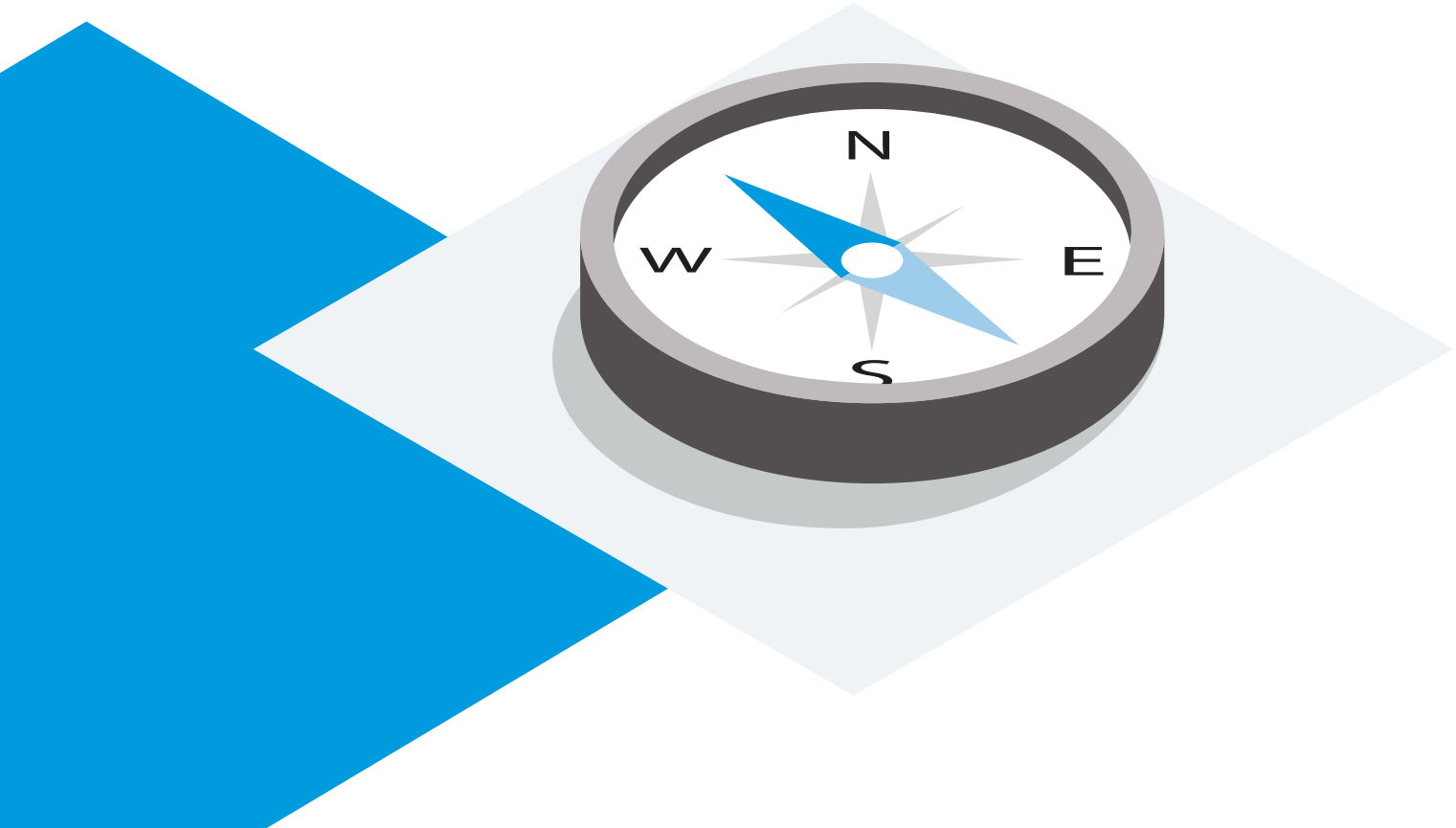


Table 1. Examples of behavioural concepts and biases

Behavioural concept or bias	Example	Implication for Explore work
Intentions-action gap - People may intend to do something, but fail to do it because of some situational factor.[11]	Someone might say that they want to eat more healthily. However, after a long day at work, they find that that chocolate bar by the checkout becomes irresistible.	Do not just rely on people's description of their intentions and assume that it will predict their behaviour. Use more objective data, like the receipts from their shopping trips, to validate behaviours or ask if you can observe them when they are shopping.[12]
Dual systems theory - We have two types of thinking: one fast and intuitive, the other slow and deliberative.[13]	People are more careful in their decisions when focused on some tasks, and less so on others. For example, someone might be more deliberative when searching for work, but less so when choosing a pension fund for their new job.	We should note that certain patterns of behaviour should not be attributed to a person's personality but the type of decision they are undertaking.
Attribution bias - We make systematic errors when people evaluate or try to find reasons for their own and others' behaviors.[14]	A teacher may assume that a student is lazy because they have not completed their homework, but really this could be because the student is having issues at home.	We should not assume that people's stated reasons for undertaking a behaviour are necessarily a complete explanation for their behaviour. Researchers should also take care to avoid attribution bias when interpreting their own observations. Using a mixed methods approach that focuses on the person and situation can help here.
Confirmation bias - We will tend to seek out information that confirms our prior beliefs, rather than disproves them — even if we have been instructed to disprove them.[15]	A doctor might order tests that confirm their initial diagnosis of a patient, rather than ordering a test that would rule it out.	We should try to test our hypotheses using administrative data analyses, rather than seek to confirm our prior beliefs.
Framing - The way in which we make decisions is biased by the way in which options are presented to us.[16]	People are more likely to support a policy decision that highlights the benefits (i.e. 99% of people will not be infected) rather than the costs (i.e. 1% of people will be infected).	People may respond to your survey questions differently depending on how you frame your questions, so either use one framing device consistently or present the information using both frames at all times.

Some examples of our previous Explore work

Mapping and unpacking behaviour in practice

Interviews and focus groups

- To improve successful transitions from the armed forces to civilian life, we interviewed the families of veterans and serving persons to understand the barriers to accessing and receiving support.
- To help job seekers find work, we interviewed people who had recently been made redundant so we could understand and empathise with how they experienced the process.
- To increase birth registration in Guatemala, we interviewed new mothers attending the national registration office about their experience of the process.
- To understand how to encourage innovation in small and medium sized businesses, we interviewed key decision-makers about what prompts innovation and investment in R&D.

Surveys

- To understand how people would respond to various public health communications in response to Covid -19, we ran over 80 large scale surveys and online experiments across the globe.
- To encourage healthy eating, we conducted a large-scale citizens jury to identify recommendations for government and industry.
- To test the impact of police body-worn video cameras, we surveyed officers who use the cameras when attending 'stop and search' and domestic abuse incidents.
- To encourage the ethical development of teenagers, we conducted a forum with young people and experts from the education sector, academia and industry.

Observe and Participate

- To understand how different health and social care professionals interact, we sat in on multi-disciplinary team meetings where nurses, therapists, social workers and others plan how to respond to complex cases.
- To understand the health and safety risks faced by gig economy workers, we signed up to become food delivery workers ourselves and completed several food delivery shifts.
- To understand how teachers engaged with a new program to reduce the use of corporal punishment in Tanzanian refugee camps, we prototyped early version of the program with teachers in the camps.
- To try to understand how features of online gambling sites might be encouraging problem gambling, we spent time gambling with real money across a number of games and platforms.

Administrative data analysis

- To encourage flexible working in a large government department, we analysed turnstile data to see how the entry and exit times of employees correlated with those of their managers.
- To reduce mobile phone theft, we analysed police data to determine the percentage of thefts by handset brand and relative market share.
- To improve the targeting of micro credit, we analysed a bank's customer data to identify the characteristics of firms that are most likely to pay back their loan on time, deposit savings, and graduate to a larger loan.
- To help social-workers make better decisions regarding at-risk children, we used a machine learning algorithm to identify cases which had a high risk of returning to the social care system after being closed.

Contents

Introduction

Good policy making is grounded in the everyday realities of our public services. The most effective way to uncover these realities is to Explore.

Exploring is the process of understanding the environmental influences on a given behaviour and how behaviour plays out in a given system or context. We do it by seeing how services work through the eyes of users, by talking to frontline staff, and by analysing the reams of data that organisations create during their operations. At its most in-depth, we might seek to directly experience the system ourselves. Finally, Exploring involves integrating all of these data sources into a deep understanding of a system, ideally freeing us of any unhelpful prior assumptions. With our newfound understanding of the system, we can then identify solutions that might work. Put another way, would you trust a solution that had been developed with no regard to the world in which it would be used? For example, would you get into a self-driving car that was designed by someone who had never seen a busy city intersection, or let your children get on a slide that had never been tested? Neither would we.

Although in-depth Explore work can take more time, skills and resources than you might feel you have available to you, we encourage you and your team to get out and try it, or to work with experts who can do it for you. We have written this guide to show you some of the tools available and what questions you should be asking. This guide is not an instruction manual, but a compass to steer your journey. The first half of the guide covers why you should conduct Explore work and how to conduct an Explore project, and the second half goes into more detail of the individual methods in the Explore toolkit.

Part 1: Why we Explore

Since our inception, our team has done a lot of Exploring. We have observed doctor-patient interactions in UK hospitals, analysed donations data from Australian charities, become food delivery workers, interviewed FARC ex-combatants in Colombia, and much more.

This has shown us first-hand the impact that beliefs, environmental factors and context have on behaviour. It has also helped us escape our prior assumptions, opening up new avenues of thinking that have sometimes led to surprising solutions.

Time and again, Exploring, or 'getting under the hood' of a problem, has led to our most effective interventions. We do this because it is a fundamental part of making better policies, services and products. Failing to Explore the context of a behaviour change problem can often lead to ineffective solutions or can backfire. While it can be challenging, particularly when dealing with sensitive or complex social problems, it is often incredibly rewarding and inspiring. We hope you find this too.

Introducing the four Explore tools

The aim of the Explore phase of a project is to understand the context. We do this by using our Explore tools to seek out in-depth information that covers a range of people. We also want to understand 1) what people do (as observed from the outside) and 2) why they do what they do. The latter is only revealed by understanding their thoughts, attitudes, values, personality traits, intentions and/or motivations toward a certain behaviour which are not directly observable. Different tools are more efficient at providing different types of insights. The table below provides a rough guide on which type of information each tool can efficiently provide.

Table 2: The four Explore tools

Tool	Description	Ideal for	Efficient at providing
Interviews and focus groups	Speaking with people to investigate their views, experiences, values, emotions and motivations.	Researching complex issues that require in-depth understanding, or sensitive issues that require delicate handling.	In-depth information about what people say they do and why they say they do it.
Observe and/or participate	Doing or seeing something yourself, for example signing up for a government service, or watching how a service is delivered.	Trying to understand a culture, process or service, or examining people's behaviour and interactions.	In-depth information about people's observed behaviour.
Surveys	Surveying a large number of people to get an understanding of how an issue affects them.	Contrasting and comparing different people's perspectives and experiences.	Information about what a range of people say they do and why they say they do it.
Administrative data analysis	Looking at data to find relationships, patterns, and trends.	Getting an understanding of an issue from a large sample and understanding what people are doing from the system's point of view.	Information about the range of people's observed behaviour.

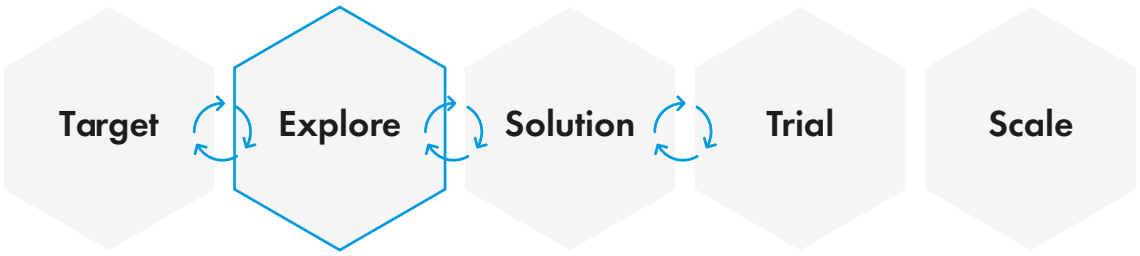
The four Explore tools are drawn from a variety of academic fields, ranging from psychology and economics, to anthropology and design. As they require more than a handful of pages to fully detail their application, this guide is not designed as a comprehensive 'how to'. Instead, it is a way to quickly equip yourself with the basics, and to begin developing the confidence to use them effectively.

When should you do your Explore work?

Explore is the second stage of our project methodology, TESTS, which stands for:

- **Target:** Identify and prioritise the specific behaviours to change.
- **Explore:** Understand the context of these behaviours.
- **Solution:** Design your solution.
- **Trial:** Test your solution to see if it works.
- **Scale:** Find ways of bringing a successful solution to new contexts

TESTS has many parallels with what is often referred to as the double diamond design process developed by the British Design Council. It involves opening your mind to a range of ideas, before tightening your focus as you seek, tune and refine solutions to trial and scale up.



Box 1. Expanding your thinking

Trapped by our existing knowledge and assumptions, we can often find it hard to see the world in new ways. We will often search out information that confirms our own worldview and dismiss information that disconfirms it.[17] These two processes are both forms of confirmation bias.[18] Sadly, you cannot outsmart confirmation bias. In fact, it can have a larger effect on people who are more numerate.[19]

When conducting Explore work it is important to avoid confining our worldview or dismissing information that disconfirms it. We want to understand the world as is and avoid biasing our perceptions of it with our prior conceptions.[20] Thankfully, human beings are better at avoiding confirmation bias when they think about issues using concrete, rather than abstract, examples.[21] As Explore work involves understanding the day-to-day realities of an issue, it might also make it easier to avoid confirmation bias.

To avoid confirmation bias after you have completed your Explore work ask yourself: how would you respond if your Explore research had given you the opposite results?[22]

6 key tips for successful Explore work

From our experience, these tips will help ensure that your Exploring will lead to practical insights:

1. **Work out when you need to use the cartographer or surveyor mindset**
2. **Remember that different data sources provide different information**
3. **Don't go it alone**
4. **Be prepared for ethical quandaries**
5. **Conduct a pre-mortem to identify mistakes before you make them**
6. **Iterate, iterate, iterate… ….and evaluate**

We'll describe the trials and tribulations underpinning these lessons below.

1. Work out when you need to use the cartographer or surveyor mindset

There are often two reasons for Exploring. The first is to try to understand the policy context as it is, and the second is to understand whether there are components of the context that you can change to encourage a certain behaviour. In the case of the former, the focus is on understanding the current state of the context and is more descriptive and backwards looking. This focus requires you to examine the past or the present and ask: **'what did/does happen?'**. You should think of yourself like a photographer or cartographer, documenting the world.

The second reason for Exploring is because you are trying to make more inferences and predictions about how your intervention (be it a new policy, service or product) interacts with the context. This requires you to ask: **'what could/would happen if this change was made?'**. Here you should think of yourself like a surveyor, who takes note of the local environment to decide whether or not the environment will support a hypothetical structure that will be built and consider whether it will serve the people who will eventually live there. They may also consider how the local environment might be impacted by those changes. They often do this without a fully formed notion of what will be built.

In most behavioural projects, you will probably start by focusing on what the current context looks like, and then shift focus to what we can do to change the context. This is usually a subtle and iterative shift, rather than a clear one. For example, imagine that you have just been put in charge of a project to stop social housing tenants falling into arrears — that is, falling behind on their rent. At first, you should try to understand why it is that tenants fall into arrears (**'what did/does happen?'**). Some tenants may fall into arrears because an issue with their property has not been fixed by the housing provider. While the provider might see this as two separate issues, they are linked in the minds of the tenants. These tenants feel they should only pay rent for a properly maintained property. If the provider does not uphold their side of the deal, then why should they pay rent?

Once you have uncovered this finding, you should switch your focus to understanding whether introducing an intervention (for example, faster repairs) would affect rental arrears. This moves you from the cartographer mindset (mapping the world as it is) to the surveyor mindset ('if I were to introduce this, what could/would happen?'). If you are suitably convinced that this is the part of the problem to address, you then need to understand more about the process that the provider currently uses to schedule repairs. In this case, you need to shift back into a 'what did/does happen?' state of mind. Once you have understood this process, you then need to think about which parts of the maintenance process you can tweak — making you a surveyor once more.

2. Remember that different data sources provide different information

Within government, quantitative and qualitative research is often done by different teams or commissioned from external organisations. It is also often conducted and analysed in isolation, rather than combined to create a holistic understanding of a problem. Government economists, for example, spend a lot of time analysing data, but rarely speak to frontline staff. Similarly, although many governments now recognise the benefits of qualitative inquiry, it is often conducted separately from any quantitative data analysis. The findings from either are rarely triangulated.

Where possible, we recommend using both qualitative and quantitative tools, since they are complementary. Their combined value is far greater than that of the sum of their parts. Comparing and contrasting findings from multiple data sources also enables triangulation – that is, it can reinforce, contradict, or add nuance to your insight. We have regularly found these comparisons to be an invaluable part of the process.

In the academic world, triangulating qualitative and quantitative data was seen as controversial.[23] This debate has largely focused on whether or not data from qualitative and quantitative methods can be used to verify each other. The broad modern consensus is that these methods offer different dimensions of the same phenomenon. Before you start your Explore work, you should consider whether the method you are using is the right tool for the dimension you are looking at.[24]

3. Don't go it alone

Wherever possible, you should co-create your Explore plan and test your findings (and, ultimately, design your solutions) with key stakeholders outside your project team. Problems are rarely novel and it is likely that many people before you have tried to solve the issue you are looking at, or one analogous to it, many times over. By speaking to others, you are increasing the mental bandwidth being used to solve the problem you are looking at, while also harnessing the many hours of work they have previously dedicated to solving the issue.

When sharing your insights with others, consider how you frame sensitive issues and talk about the people participating in your research. For example, you should never forget that you are describing human beings with agency, rights, values and emotions. You should also try to make these sessions as engaging as possible through the use of visuals and tools. For example, you can use:

- **Journey or process maps** – Creating a detailed map from your findings is a great way to communicate what you have learned, and to identify the most promising points for change. Through Exploring, we have created journey and process maps for everything from breast cancer screenings to understanding why people litter. These are more useful when you are in the 'cartographer' mindset, trying to work out what is happening (see point 1).

- **Personas** – Personas are fictional characters that are used to represent certain user types, developed by software designer Alan Cooper.[25] They can be useful for guiding stakeholders to consider the full range and diversity of users. Unfortunately, however, they are often done poorly and only highlight the most memorable or dramatic cases or are so specific that policymakers end up developing a policy for five fictional individuals rather than the whole population. (We provide an example of this on page 25). For this reason, it is important to make sure they are grounded in quantitative or qualitative data,[26] not just your opinions or beliefs, or the opinions or beliefs of your stakeholders.

- **Prototypes** — When you are in the 'surveyor' mindset, you may have some early ideas of what your Solutions might look like. Rather than investing heavily in a single option, it can be useful to develop early prototypes of these. If you are looking at developing a website or a piece of software, wireframes can be really useful to understand whether there are any pain points in the user interface. If you are developing a new service that is delivered by frontline staff, then a role play or a table read of a new script can be helpful to gather first impressions of specific phrases. We cover these in more detail on page 31 and 37.

- **Systems maps** — These are used to identify feedback loops in systems that affect behaviours. Importantly, they can be used to identify leverage points within a complex system where a small shift in behaviours (potentially facilitated by a nudge) can produce big changes. Donella Meadows' 'Thinking in Systems'[27] is an excellent primer here. However, systems maps are not without their downsides (see Box 2. Systems maps and choice overload).

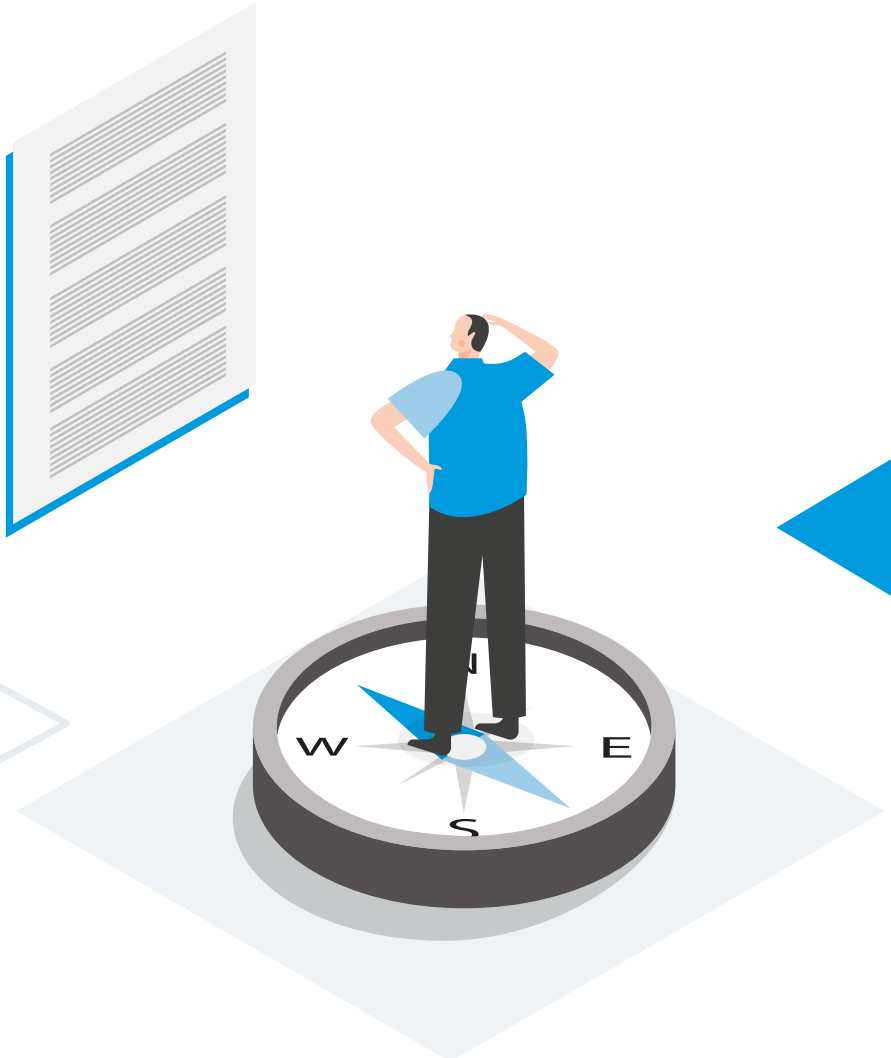
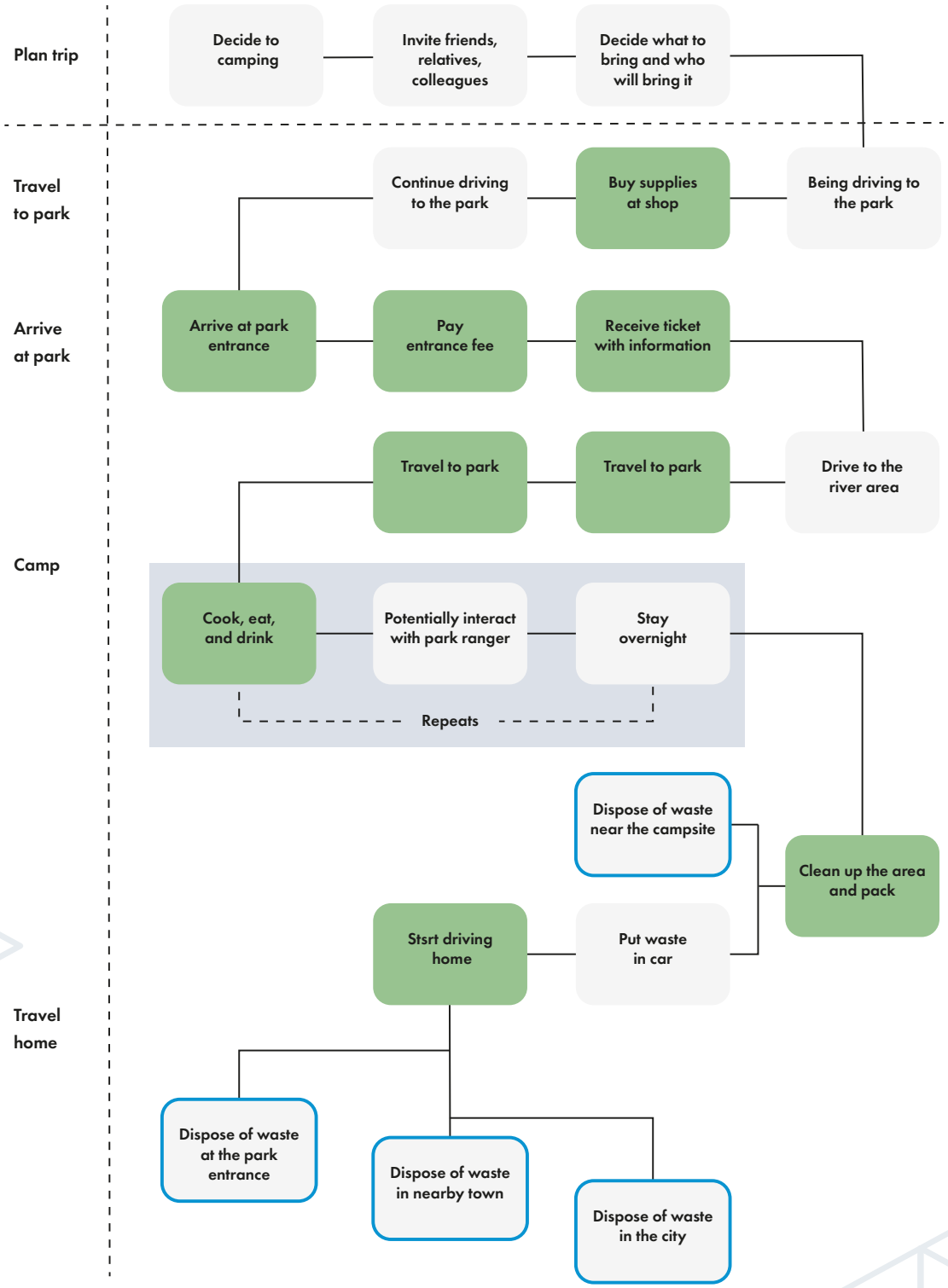


Figure 1: We created this journey map for a project with the UNDP Mongolia Country Office that was aimed at improving waste collection at Gorkhi-Terelj National Park



KEY:
Blue = Target behaviours
Green = Steps in the process where a solution could be implemented
Grey = Steps in the process where a solution could not be implemented

Box 2. Systems maps and choice overload

Many people will often develop systems maps with the aim of comprehensively describing all the causes of a complex policy problem, like obesity. This can be a useful exercise in the 'cartographer' phase of your project (see page 12), because it will allow you to identify causal factors that you might not otherwise have appreciated. However, mapping out too many causal factors can give you a choice overload[28] when it comes to choosing a particular issue to focus on. It may simply discourage you by giving the impression that the problem is too complicated to address easily and there is nowhere obvious to start.

4. Be prepared for ethical quandaries

It is important to conduct research to a high ethical standard, and you should do an ethical review to ensure you have identified anything that may harm your participants or those carrying out the research.

The level of detail and external scrutiny needed here will depend on the nature of the issue you are researching and the groups you are dealing with. For example, if speaking with children or vulnerable groups, consult your home institution or government department's approach to ethical oversight. This may be an internal ethics review board, the ethics review board of another department, or an ethics review committee of a University (See box 3 for more information).

Box 3. Working with sensitive topics or groups

When Exploring, you should always make sure that you conduct your research ethically and understand any sensitivities that relate to the issue you are investigating. Recognising the relevant sensitivities can be especially difficult if you do not have much experience with the issue that you are Exploring, or you do not have a connection to the people that you are hoping to speak to. For example, asking direct questions may not be considered a faux pas in your cultural context, but this might be the case in a different culture.[29] You should always treat your participants with respect. You should think about the nature of the topic you hope to Explore, what information might be unintentionally revealed through the Explore work, and the needs and values of the people that you are speaking to.

Once you have done this, you should think about the processes that you put in place to make sure that any concerns are addressed. The first consideration will usually be recruitment and consent, then the risk mitigation processes you put in place during the Explore work, and finally, what happens at the end of the project. All of these should be decided before you start your Explore work, but could change as the work progresses, as you might identify new risks.

When it comes to recruitment, you should think about the way in which you approach potential participants, including the processes you use to obtain their contact details. Once you have decided this, you should think about the information you give them in the consent process (or whether you can use an opt-out consent process or waive it entirely). Any information given in the consent process should be enough to give the individuals an understanding of what they are signing up to and clear enough so that you can reasonably assume that someone has understood it. If your consent form looks like a technology company's End User Licence Agreement, you should start again. The exact information that needs to go into a consent form will vary depending on what you are doing and where you are doing it, so make sure that you look this up.

In terms of the risk mitigation process, you should consider all of the potential risks that might occur (a pre-mortem can help here — see below) and then try to mitigate them. These risks might include: a breach in confidentiality by the participants or the researchers, the disclosure of illegal activities (in which case you need to be clear on the obligations of the researchers and the implications for the participants), whether you are going to use a debriefing process, and whether the way in which the research is done might put the researchers at risk (for example, if you are conducting research with domestic violence defendants, how can you ensure that the researchers are safe)?

Finally, you should consider what happens to the information that is collected. Think about whether individuals are named or given a copy of the results (which can be difficult if you have purposefully avoided collecting contact details). However, you should also consider how the data is stored, when it is deleted and who should be deleting it. You should also consider the social licence that you have when using any data.[30] This relatively new concept considers whether or not the public and any affected groups would accept the use of their data in a certain manner. Work in New Zealand has been especially progressive in this regard and the guidelines on trusted data use are worth reading.

5. Conduct a pre-mortem to identify mistakes before you make them

We are often overly optimistic about our plans. We underestimate the time tasks will take, and do not build in a buffer to overcome setbacks and chase leads.[31] This bias, which is known as the 'optimism bias', can have a big impact.[32] It can push out your timeline, increase costs and derail your project.

To minimise its effects, once you have a plan for your project, but before you launch it, you can do a 'pre-mortem'.[33] As organisations and people approach a decision, they get locked into it, making it difficult for team members to speak up when they identify an issue. Pre-mortems, however, legitimise dissent. They are very simple to run:

1. The project team and other relevant stakeholders get together in a room – Having outsiders to the projects in the pre-mortem helps avoid groupthink.
2. The team leader sets the scene – the team members are in the future and the project has been a spectacular failure.
3. Everyone in the room independently writes down reasons for the failure.
4. The team discusses these reasons and revisits the plan.

Pre-mortems are extremely helpful because they turn finding flaws into a positive process, help to minimise harmful groupthink[34], and give more junior staff a safe space to voice their opinions. They can also help you to develop mitigation strategies in advance.

6. Iterate, iterate, iterate... and evaluate

This tip is taken directly from the human centred design playbook.[35] We cannot emphasise the value of iterating your approach based on your learnings enough. This requires a high tolerance of ambiguity, which can initially be frustrating. You may feel like you are not making much progress or that everything that you are doing is only uncovering more questions, rather than answers.

You can avoid the feeling of a lack of progress by setting out which key questions you need to answer through your Explore work. These questions will become your learning milestones. The notion of a learning milestone has become a popular method among startups to document progress during an information gathering process.[36] Learning milestones are often described as the questions that need to be answered for the project to move forward.[37] These might be: 'which specific behaviour should we target our intervention at?' and 'which pain point in this process should we focus on?'. By setting these questions out, you can then evaluate your progress against these milestones. It is also a good idea to have others evaluate your progress against your milestones through discussion or user testing.

Progress against these milestones is not irreversible. There are many reasons to go back to a question which you thought you had previously answered. However, specifying these milestones in advance can help you maintain focus and motivation.

Planning to Explore

When Exploring, preparation is key. Once you start exploring, your plans can (and likely will) change. This is expected; however, consciously adjusting a plan is always better than Exploring on the fly. If you do not have a clear sense of what you initially want to find out, then you risk scope creep and/or conducting a number of activities that ultimately do very little to address your original brief. When you start your preparations, it is important to:

1. **Understand what you already 'know'**: It is rare that you go into an Explore work without any preconceptions about the context. Some of these will help you, but some might lead to unhelpful assumptions. Before you start any Explore work, it is worth explicitly articulating what you 'know' about the context. This might be on the basis of previous conversations that you have had with stakeholders or evidence reviews. This does not have to be a large investment of time. Even a one hour meeting with your team will help get everyone on the same page and identify blindspots, assumptions, differences in terminology and misconceptions. (Noting that longer and more complex projects would benefit from a longer meeting).

2. **Understand what you want to find out:** This is the most important step for any Explore work.[38] You should articulate what you hope to learn and why you want to learn it (remember the distinction between the cartographer and the surveyor from the previous section). Your questions might include one or two broad overarching questions and a number of more focused questions that sit underneath these. You might iterate or refine these throughout the Explore process, but it is important to have a broad sense of what you want to find out before you start.

Note that these research questions differ from specific questions that you might include in a survey or an interview. They focus instead on the aims of the project. They will inform everything you do, from writing interview questions to data analysis. An example of an Explore question might be: 'What can be changed in employment centres to improve employment outcomes?'. Note that there are lots of sub-components to this question. For example:

- What are the processes in employment centres?
- Which of these can be changed?
- Which of these are effective in improving outcomes?

Each one of these questions can then be broken down even further. There is no clear cut-off for when you should stop breaking your questions down. However, when you realise that the resources that you have at your disposal would never allow you to fully answer the questions you've written out, you should stop. This is more of a risk when you are in the cartographer mindset. You should also stop if your questions are so detailed that you are focused on nuances that are unlikely to make a difference to your solution. This is more of a risk when you are in the surveyor mindset.

3. **Understand what you have at your disposal:** Review your problem,resources, timelines and where your stakeholders stand to ensure your project has a clear purpose, and you have the buy-in, people, and budget to do it.

4. **Draw up a plan:** Once you have worked out your questions and taken stock of what you have to answer them, you should create a plan to see how you will use your resources to answer your questions. It should include:

   - **What?** The combination of tools and techniques you will use to answer your Explore questions.
   - **Who?** The groups, organisations or individuals you will speak to, observe, survey or collect data from. (See Box 4 below for more details)
   - **When?** Your timeframe and schedule.
   - **Where?** The locations where you will conduct your exploration.
   - **How?** The way you will carry out your exploration, including how you present and introduce yourself, and how your Explore questions and chosen tools will help answer them (for example, specific interview and survey questions).
   - **Why?** The rationale for using the specific methodology you have chosen.

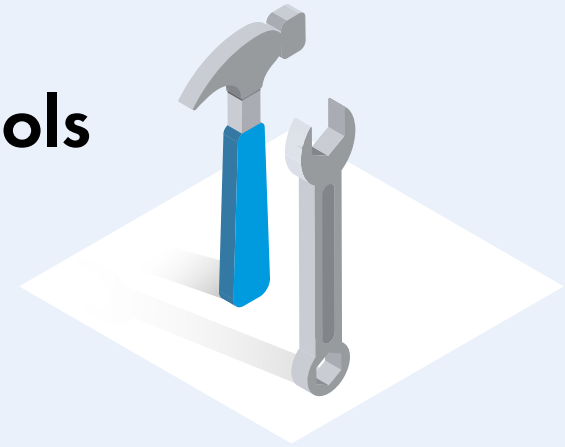
We have developed a field guide to help you plan out our Explore work. You can access it on the Behavioural Insights Team website.

Box 4. Whose behaviour and attitudes should you be Exploring?

Who you should speak to will vary greatly depending on what you want to understand. In many quantitative analyses, the aim of the research is to get 'representative samples'. However, this is not always the case. Here are some different approaches to sampling that you might want to consider.

Type	Description	Why use them?
Homogeneous sample	People in the same group with the same characteristics	You only care about the behaviour of a very specific group.
Stratified sample	People who are fairly homogeneous, but vary in a particular dimension. For example, the school year.	You care about a specific group, but want variance within that group.
Deviant sample	People who are extreme or unusual cases. For example, people who are very good or very bad at their jobs.	By looking at people who are exceptionally good/bad at what they do, you might learn of specific effective/poor practices.
Heterogeneous sample	People who differ from each other on an important attribute. This might be a specific behaviour of interest.	You want to understand the range of behaviours that people show.
Typical cases	People who are deemed 'normal' or 'average' in a specific context.	Even asking someone to provide a 'normal' sample will tell you a lot about the way in which others see the system.
Convenience sample	People who you can get hold of.	Recruitment can be difficult and speaking to someone is usually better than speaking to no one (as long as you understand that their views may not be representative).

Part 2: The 4 Explore tools



We will cover 4 Explore tools in this section. These tools are:

- Interviews and focus groups
- Observation and participation
- Surveys
- Administrative data analysis

Why these tools?

These tools may not seem like the most exciting methods available to would-be Explorers. However, every Explorer should have some familiarity with these fundamentals before moving to more complex methods. These are the tools that we most commonly use and have a set of strengths and weaknesses that complement each other nicely. At the end of each section, we have given examples of more complex versions of these tools.

What are the relative advantages of each tool?

The four tools broadly cover all four quadrants in the matrix on the next page. On one axis we have breadth and depth. This is because some of the tools allow you to obtain deep insights (usually from a smaller group of people) and the others allow you to capture broad insights from a larger group of people.

The other axis divides tools that focus on people's internal thoughts from those that focus on their observable behaviour. For example, surveys and interviews are best used to capture what people think about a behaviour, and you can ask them to reflect on why they are doing something (their thoughts, attitudes and beliefs). However, big data analyses of administrative data and observations are generally more effective at capturing what people are actually doing (their actions, the triggers of their actions and the impact of their actions).

This distinction will give you a framework to work in. Ideally in any Explore project you will want to understand issues in depth, but also get some breadth. You will also need to see what people are actually doing and understand why they are doing it. This framework is designed to give you some basic guidelines. Some tools can cover two quadrants. If you are conducting a 'service safari'[39] in which you become an active participant in the process you are studying, you will be able to record what you and others are actually doing, whilst monitoring your own internal thoughts about why. As long as you remember that you should aim to cover all four parts of the framework, you will get some valuable insights.

Which tool to start with?

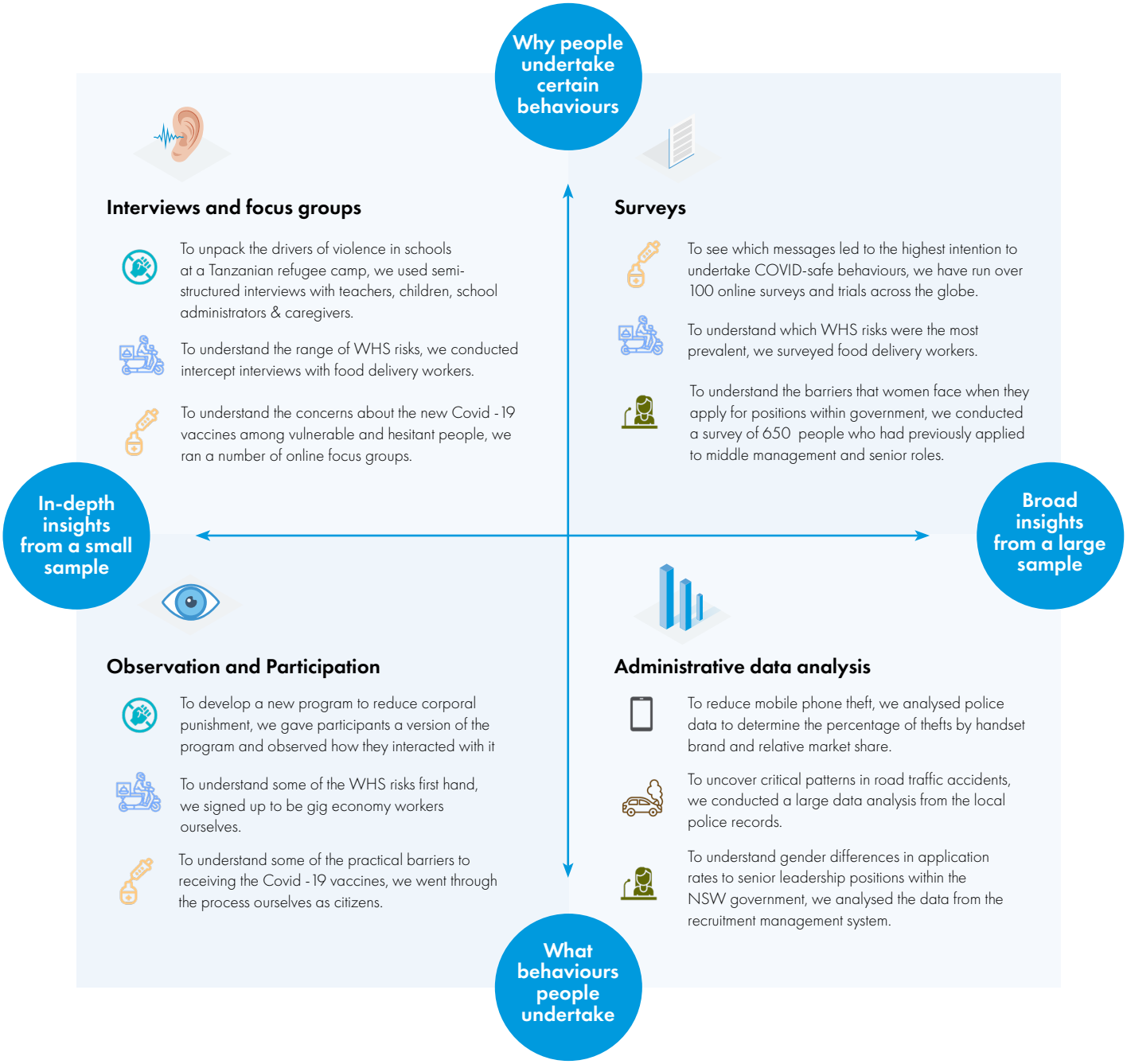
There is no set order for which tool you should use first. In some cases, you may have access to administrative data before you have time to develop any interview guides. Administrative data can also be useful to guide later activities. It can help you identify high or low performing sites (which you can visit to conduct observations) or individuals (who you can interview). If you are choosing whether to use interviews or surveys first, this will really depend on your research questions. Surveys generally require very precisely constructed research questions, whereas interviews are more forgiving of less defined research questions.

If you know very little about the types of attitudes that people might hold, then interviews are a good place to start. You can use them to identify the range of attitudes that people might hold. If you survey a group of people in your population of interest, you can then see how prevalent some of these views are.[40]

However, it might be the case that you already have a sense of the range of views that people might hold and have a precise set of research questions. In this case, you can start with the surveys and then use interviews to unpack people's responses. There are also logistical reasons for starting with a survey. You can use a mass mailout survey as a way to recruit participants for later interviews. Whichever path you choose, you may still need to iterate your plans based on your findings.

Mapping the relative advantage of each Explore Tools

The diagram below shows how each tool has its own relative advantage in capturing depth vs breadth and in capturing people's internal thoughts or observing their outward behaviour. Any tool can be used to capture information in another quadrant, but it is generally not that efficient at doing it.



Key

- Understanding the Work Health and Safety (WHS) risks faced by gig economy workers (see page 40)
- Developing effective public health messaging for Covid -19 (see page 49)
- Reducing violence in Tanzanian refugee camps (see page 32)
- Understanding road traffic risks in the UK (see case study on page 6)
- Developing the Mobile Theft Index (see case studies on page 6)
- Encouraging female senior leaders in the New South Wales Government with the Department of Premier and Cabinet. (see case studies on page 6)[41]

Interviews and focus groups

You may have been an interviewer or have been interviewed before as part of a recruitment process. Therefore, it is tempting to use the same methods for Explore work. However, there is a key difference between recruitment and Explore interviews. When interviewing to recruit, you are assessing and judging. In an Explore interview, you are listening and learning to gather information about a process or system, and investigating a person's views, experiences and motivations.

One of the benefits of interviews and focus groups not found in other Explore tools is the opportunity to ask follow-up questions. You can unpack answers, probe interesting leads, and deep dive into an issue to understand the system, people and processes involved.

Importantly, although they can seem easy, interviews and focus groups require considerable skill and preparation. For example, to reduce the effect of behavioural biases, you need to consider how best to structure your questions, how to dress, and build rapport. Otherwise, your interviewee may not feel comfortable being open and honest with you.

Step 1: Choose your format

We will discuss interviews and focus groups together in this section, and have used the term 'interview' to describe an interview in which you have one participant and 'focus group' to describe any interview with more than one participant. We will cover them together because you will need to make similar design choices if you run interviews or focus groups. Both can be used to understand the views of participants, but have different strengths and weaknesses. Interviews are best used when you want to get deeper insights from individuals, because they allow you to follow up on specific points that your interviewee makes.[42] If your subject matter is sensitive in nature, then people are also more likely to open up in one-on-one interviews than in larger groups (see box 5). However, they can be expensive to run and might hide some social dynamics that you wish to observe.

If you want to understand how people interact with each other or you want to establish where there is consensus and where there is disagreement, then a focus group can be helpful.[43] Focus groups are more than a collection of individual interviews and may not capture some of the depth of individual interviews. However, they allow you to observe group dynamics (which are used to shape the discussion).[44] These group dynamics could either be between strangers or between specific participants (for example, couples discussing their household budget or an apprentice and their employer).

Box 5. Understanding Social Desirability Bias

In interviews, we rely on people to give us honest and accurate reports of their experiences, behaviour, decisions and motivations. However, deliberately or otherwise, they may not.

One reason this can happen is we like others to see us in a positive light. This is known as social desirability bias. Social desirability is one of many response biases and was first identified in personality research when looking at people's responses to surveys[45] (we cover techniques to overcome this in the next section). It can also manifest in interviews. In fact, there is evidence that it is greater in face-to-face interviews as people feel more anonymous when they are not giving their responses directly to a human being.[46]

The social desirability bias is one form of response bias. Other response biases can cause the person you are interviewing to over-report 'good' behaviours (for example, time spent exercising or reading to their children) and under-report 'bad' behaviours (for example,caloric or alcohol intake),[47] or to give us answers they think we want to hear because they 'don't want to ruin the research' (also known as 'demand characteristics').[48] In some instances, young people may overreport 'bad' behaviours to cultivate a certain image.[49] People may also give responses because they do not want to be rude to others, which is sometimes referred to as courtesy bias.[50]

Carefully wording your questions and prototyping them with your team is your first line of defence against these biases. If you are trying to understand a specific phenomenon, then asking people for illustrative examples can help you overcome response bias. For example, if you ask a student whether they like an online teaching module, they may feel obliged to say 'yes'. A follow up prompt on what they liked about it may uncover that their endorsement was not wholehearted.

You should also be aware of the context in which you are asking. If you are asking a nurse about their views about a hospital discharge process with a senior representative from the Department of Health present, they might be less likely to voice their complaints about the process. This can be true even if the Department of Health representative was not involved in the design of the discharge process. They might just be seen as an authority figure. Power dynamics can have a big impact on response bias and should be factored into the design of the interview.

Step 2: Identify your participants

Before you can begin your interviews, you need to choose who to interview to help you answer your Explore questions. The Planning to Explore section covers how you can do this. When working out who you should speak to, we would recommend investigating a diversity of experiences and views. Interviewing 100 people with the same background can give you less information than 10 with different backgrounds. What you learn from a diverse range of interviews can stretch your thinking and lead to more innovative solutions. But be careful of overly focusing on outliers — the availability heuristic can mean that memorable anecdotes might seem more common than they really are (see Box 6, below).

Box 6. The Availability Heuristic, Vividness and Personas

The availability heuristic was first described in 1973, by Amos Tversky and Daniel Kahneman.[51] They found that people thought that things that are easier to recall (i.e. words beginning with K) are more common than things that are harder to recall (i.e. words with K as their third letter), even if the things that were harder to recall were actually more common.

This heuristic can affect people's judgement of how big an issue is. For example, we worked with an agency to reduce the number of social housing tenants in rental arrears. As part of this project, we spoke to tenancy officers about the number of tenants that were in arrears. A large part of tenancy officers' roles was to follow up with tenants to make sure that they paid their rent. When we asked these tenancy officers what proportion of tenants were in arrears at any given time, some officers' estimates were nearly ten times higher than the true figure. This is because a large part of their day involved speaking to the tenants that were in arrears and therefore they assumed that this segment was much larger than it actually was.

Later academic work has found that more 'vivid' examples are also subject to this availability heuristic.[52] Therefore, extreme examples might be seen as being more common than they actually are. A few years ago, we ran into this exact issue when we were working with an organisation which was trying to get people back to work. We were not able to conduct interviews directly with the injured workers, but spent a long time working with the case managers to understand the people that they were helping back into work.

We started one workshop by developing personas in groups. Personas are fictional characters created to represent a user type, which are widely used in marketing and user research. The groups came up with four personas. The first three vividly described people who were difficult to help. They were suspicious of the case managers and the wider system. They were withdrawn or short tempered. However, the fourth persona was much easier to communicate with, more proactive and more pleasant to help. When we asked the staff how many people on their caseload were like the fourth persona, the group pretty unanimously said '80%'. Had we not sense-checked how representative these personas were, we would have designed a solution assuming that most people were much harder to help than they actually were.

At this point, you should also consider how many interviews or focus groups you want to conduct. The question of how many participants you should speak to is a tricky one. The number of people you need will depend on many factors, so we have included these four rules of thumb as a rough guide. You will need a large sample, or to speak to more people, if:

- Your research question is multi-faceted.
- Your population of interest has a diverse range of characteristics.
- There is little pre-existing literature on the topic.
- You want to be able to compare across groups.
-

Both interviews and focus groups can be time-consuming, and generally speaking, you should continue interviewing until you are hearing consistent stories and are not uncovering new insights. In the academic jargon, this is referred to as 'thematic saturation'.

Thematic saturation can often be reached relatively quickly if you have ensured that the scope of your questions is tight and if your participants have similar experiences with the phenomenon under investigation.[53] Despite this, it is not always possible to conduct enough interviews to reach 'thematic saturation'. This can be for time or budgetary reasons. If you do think that you will not reach thematic saturation, you should try to rescope the focus of your interviews. If this is not possible, you should take note that the responses that you have gathered may not cover the full range of views, experiences or attitudes of your population of interest and ensure that your Solutions are designed accordingly.

Step 3: Write your interview questions

Before you start writing your questions, you should decide how tightly you want to script your interviews. Most of the interviews or focus groups we do use a semi-structured format, in which you have pre-written questions to draw from, but you do not let them entirely constrain the interview. They allow you to ask follow-up questions to unpack the answers you receive. In our experience, semi-structured interviews provide helpful consistency along with the opportunity to uncover personal insights.

To write your interview questions, you should first refer to your Explore questions. They will provide direction and will ensure that the interview questions you write will help you uncover the information you need. The number of questions you write will depend on the length of your interviews or focus groups.

When conducting semi-structured interviews, we recommend aiming for a length of around 45 minutes, or within a range of 25-60 minutes.[54] This will allow you to deep dive into your topic, while respecting your interviewees' time. That said, the length of your interview will depend on the availability of your participants and the scope of the interview.[55] Regardless of the total time of your interview, you should make sure you leave plenty of time for you to build rapport with your interviewee at the start, and end with some time for them to add anything that your interview script hasn't covered. The last 5 minutes of any interview can be surprisingly useful when the more formal Q&A approach of an interview ends.

Focus groups tend to last slightly longer, at about 1.5 to 2 hours.[56] To get the most out of focus groups, you will need a skilled moderator to steer the conversation, to keep participants on topic and to encourage everyone to speak up. We would also recommend having a note taker in the session to ensure that all insights are covered.[57] They will be able to keep an eye on whether or not a false consensus is being created because group members are not voicing their honest opinions in order to create consensus and minimise conflict.[58] Ensuring that you have a diversity of opinions in the room can help overcome groupthink,[59] as can asking people to write down their thoughts before discussing them as a group.[60]

Importantly, resist the temptation to use your Explore questions as your interview questions. Doing this will result in overly direct questions, and may mean that you are:

- steering participants towards your own hypothesis;
- asking participants to make inferences about what is driving the behaviour of others, which is something they may not be well placed to do; and
- narrowing the focus, when you should be trying to step back to see the wider picture.

In addition, think about the order in which you ask your questions, as this can affect the tone and flow of your interview. To quickly build rapport, we recommend starting with broad questions, before building to specific, personal ones. For instance, when we were looking at ways in which we can encourage women to attend breast cancer screening appointments, we started with questions that were focused on logistics — for example, where and when they were last screened, and how often they get it done. Questions like these provide a platform for then asking other, more personal questions. An example of how you might do this in a 45-minute interview is shown in the table below.

Table 3: The outline of a typical semi-structured interview

Section and Timing	Purpose
5-min Introduction	Outline the purpose of the interview or focus group and set the tone. For example, by stressing that there are no right or wrong answers and encouraging the participant to be open and honest. If you have not already obtained consent from your participants, you should do so here. If you have already obtained consent, it is worth confirming it here.
10-min Warm up	Build rapport through common areas of interest by asking simple, introductory questions. For example, if you are interviewing frontline staff, you could ask them to describe their role and what they like best about it.
15-20 min Deep dive	Ask deeper and more difficult questions, seek specifics, and probe interesting avenues of conversation. A good rule of thumb is to have five key questions, which form the 'spine' of your interview. If you are asking sensitive questions, you should also try to ask them in the middle of the interview. This will allow you to return to a 'less emotionally charged' subject by the end.
10-min Wrap up	This is where you can start wrapping up the interview. You should thank the participants for their time and give them the opportunity to add anything that you haven't already covered. You can also ask them if there was something that they expected you to ask, but didn't. In the same way that the light before sunset is referred to as the 'magic hour' in Hollywood, the last few minutes of an interview will often produce some golden insights. This is because participants will have had time to reflect on their answers and synthesise their thoughts. Ending an interview too abruptly will mean that you miss these insights. If possible, you should also ask if you can contact participants with any follow up questions. This can be useful as you may discover new questions to ask as your exploration continues.

As mentioned above, focus groups tend to take a little longer than individual interviews. They follow a similar format with a few exceptions. If the focus group is being conducted in person, you should provide snacks or refreshments for your participants. If you are leading the focus group, you should do this by setting some ground rules to ensure that the participants understand the focus of the discussion and how it will be conducted.

You should also cover the content of the consent form. After this, everyone should introduce themselves. This allows the people in the group to rehearse the two actions we need from them in the discussion: to speak and to listen. After these introductions, the facilitator should introduce the first topic to be discussed and the structure of the focus group should follow something similar to that described in the table above.

Once the focus group has started, you should make sure that everyone has the chance to speak about one of the topics. However, it is not good practice to get each person to respond to each question. It is repetitive and disrupts the flow of a focus group. Remember, your group may not be equally affected by the topics that you hope to discuss.

Box 7. Avoid these interview question pitfalls

In interviews, how questions are worded can affect how they are answered. For example, framing questions in certain ways can lead to less accurate responses. To stop this happening, avoid the following types of questions:

- **Leading questions.** For example, do not ask 'Why do many job seekers not meet their minimum job search activities?'
- **Questions containing social norm.** For example, 'Many job seekers say that they should be able to claim their benefits online. What do you think?'.
- **Closed 'yes' or 'no' questions.** For example, 'Do you think that job centre advisors should apply for jobs directly for job seekers?'.
- **Double-barrelled questions that are two questions in one.** For example,'How often do you look for work and update your CV?'.
- **Complex, difficult to understand questions.** For example,'Behavioural science provides compelling evidence that commitment devices can increase self-control and reduce the intention-action gap. Can you think of examples of this from your own life?'
- **Not considering the flow of the interview.** You should think about the flow of the interview when you are developing your questions. You might reorder the questions when you are actually running a semi-structured interview, but you should design an interview so that the user-experience of your interview does not jar. Avoid quick transitions from topics that might make someone happy to ones that might make them sad, or from neutral topics to ones that are emotionally charged.

Step 4: Conduct your interviews or focus groups

When conducting your interviews or focus groups, the first thing you need to do is begin building rapport with your interviewees. Good rapport with your interviewee(s) will make it easier to ask probing questions and get to the root cause of an issue. It is, however, not always easy to do.

There may, for example, be a power imbalance or cultural difference between you and the participant. To overcome issues like these, there are a few things you can do. You can:

- make it clear that the interview will be de-identified, and they can stop at any time;
- offer them a choice of interview locations that are neutral, convenient and where they feel relaxed, or give them the option to talk on the phone;[61] or
- dress in a way that does not unnecessarily emphasise the power differential (for example, avoid wearing a business suit to a community event).

Next, you need to begin guiding the discussion. You can now start to ask your interview questions, and probe interesting answers using techniques such as the 'five whys'[62]. This involves asking 'why?' in response to five consecutive answers, and it can be very effective, but use it sensitively as it can lead to defensive answers. Box 8 has a number of tips that you can use to dig deeper on certain questions.

Box 8. Dig deeper using these five probes

When used in a semi-structured interview, these probes can help you continue further down a line of questioning to get closer to the cause of an issue.

- **Clarification:** 'What did you mean when you said _____?'.
- **Details:** 'Can you tell me more about _____?'.
- **Variations:** 'How has _____ changed over time?'.
- **Counterfactuals:** 'What if _____?'.
- **Steering the conversation:** 'You mentioned ____ earlier, can you tell me more about that?'

However, remember that sometimes you don't need to prompt people. They are just considering their answer, so don't jump in too soon with a prompt!

Where appropriate, you can also let your interviewee lead the discussion to issues that matter to them most. This will make your interview feel more like a natural conversation and help you build further rapport. While asking questions, you should pay attention to your participant's emotions and reactions. When these do not match what they are saying, it might be worthwhile to probe deeper to find out why.[63] This can provide you with helpful information; for example, whether they are uncomfortable because of the topic, or because of something associated with it.

You should make sure that you look engaged in their responses in a neutral but supportive manner. Refrain from facial reactions. Remain neutral to everything the participant says, even if they say something that is shocking or surprising. Facial reactions can be interpreted by your interviewees as judgement, which can make your participant feel self conscious and shut down. In the case of excessive positive reactions, they may also feel like they are being rewarded for giving certain answers and therefore may only give you the responses that they feel you want. Active listening techniques are sometimes recommended in interviews. However, the use of paraphrasing (a core active listening technique) can often confuse participants or might be interpreted as you correcting the participant, so should be used carefully (if at all).[64] Finally, never interrupt your interviewee. If you think that they are going off track, wait until a natural pause in the conversation and then gently bring the conversation back to the topics in your interview guide.

Throughout your interview, if you are not recording, you should also try to take notes. If it is appropriate you may want to have someone else there to take notes. Conducting an interview requires you to focus on what the interviewee is saying, remembering which questions you want to ask, determining what the best way to ask them is, all while trying to build rapport. This is a hard task and one made harder if you are also taking notes.

At BIT, we often take shorthand notes and type up complete notes as soon as possible afterwards. This is a good way of achieving a balance between guiding the conversation and writing things down. In your notes, you should also include your interviewee's non-verbal reactions as this added context can tell you a lot. Taking notes, listening properly and formulating relevant follow up prompts is hard, so we would recommend trying to record interviews if possible. Modern technology has made this much easier, as you can use smartphones as a field recorder, or record a video call using software in your internet browser.[65] Whenever you are recording an interview, you should make sure that your interviewee has agreed to being recorded and knows how the recording will be used. Consent to being recorded should be obtained before you start your interview and you should remind your participants how you will use the recording when you turn it on at the start of the interview.[66]

Lastly, remember that silence or pauses can be an important part of a discussion. Forgetting this is one of the most common mistakes people make. It may feel awkward, particularly if you have asked a searching question or it is a sensitive topic, but resist the temptation to talk. Often your interviewee is thinking and will elaborate if given time.

Box 9. Consider cultural differences and adapt your approach

If your research specifically focuses on a culture that you are not a part of, you should consider whether you are the best person to be doing this work. It might be the case that a researcher from that culture will be better placed. It might also be the case that you can do that work alongside someone from that culture.

You should also consider how appropriate your adaptations are in the context of the interviews. Unnecessary adaptation for people from a different cultural background to yours can be perceived as insulting or 'othering'. This is because it requires an acknowledgement of differences between you and the person being interviewed. These perceived cultural differences may not be shared between you and the interviewee. For example, if you are interviewing second generation migrants in your home country and assume that their cultural identity is more closely linked to the country their parents emigrated from, this can be interpreted as you viewing them as not belonging to their (and your) home country. Excessive adaptations may also lead some people to believe that you are appropriating some cultural behaviours.

Additional Tools

When interviewing, you can do much more than just ask questions. These tools can be used to uncover how people behave and feel.

Card sort

'Card sorts' are designed to help participants think through their feelings. In them, they first write their priorities or concerns on individual cards, before ranking them in order of importance. Alternatively, if you have a good sense of the key issues at hand, you can prepare the cards before the interview – though it is still worth having a few blanks available for issues you may have missed and that subjects want to add in.

Visualisations

Mental mapping activities are designed to help you get a sense of a participant's world. In them, you ask participants to sketch out locations or relationships. For example, where they look for work, or the different people in their life that they talk to about their studies. You can use paper, photography, diaries or other forms of creative tools.

Pre-interview participant tasks

As smartphones are now so widespread, it is possible for frontline workers and end users to easily record their experiences through visual or written diaries before you interview them. These can consist of photos of everything they eat, or be a record of what they are doing, thinking or feeling at certain times of the day. You can then use their diary entries as prompts during your interview, to help them talk through their experience, and remember details that they otherwise might overlook.

Low fidelity prototyping

As you near the end of your Explore phase, you might have some ideas on what your Solution might look like. If this is the case, it can be very useful to show participants a prototype of this Solution and observe how they interact with it. These prototypes should be developed quickly and shown to participants early. This means that you don't get too attached to early ideas and can fight against the sunk cost fallacy.[67]

There are various different ways in which you can prototype Solutions. These might start with simple sketches of what your Solution looks like. This is a good starting point if your solution is something physical like a new form or a product. You can show these sketches to an interviewee and get their feedback on them. It is generally useful to have a few of these, which will allow interviewees to compare and contrast what they do and don't like about each one.

If you are designing a process, then you can sketch out a process map and get an interviewee's thoughts on it. You can also conduct role plays to identify difficult parts of a conversation, or create comicstrip-like storyboards which detail what happens over multiple interactions.

Citizens' Juries or Forums

Focus groups tend to involve roughly 6-8 people. However, in some cases you might want to gather the views of a much larger group. This might be because you want to establish the views of a nationally representative group of participants. In this case, you might want to use a citizens' jury or forum. These are closely related to deliberative polling, which is described on page 45. In a citizens' jury, a representative group is randomly selected from the general public to participate.[68] They are given balanced briefing materials about an issue, and over several days they come together to discuss it. At the end of the event, they then put forward recommendations for future actions or directions.

The participants are then asked to reach a collective view on the subject matter. This often requires intense debate, as well as a mechanism – and skilled moderation – to agree a collective position, such as voting and majority agreement to a proposal or position. Minority reports are also sometimes produced by these mechanisms.
A citizen jury usually refers to a smaller panel of 10-20 people. A citizens' forum normally refers to a larger panel of 20-100 people or more. A larger panel has the advantage of being more likely to be representative of the wider population, but also can make it much harder for the group to reach a collective view.

In partnership with VicHealth, we conducted a citizens' forum in Victoria, Australia that resulted in 20 recommendations to promote healthier eating and tackle obesity. At this event, an 80 percent threshold was used to establish collective agreement.[69] These events can also be run with young people. For example, in partnership with the Vincent Fairfax Family Foundation, we conducted a two-day forum with 61 young people aged 12 to 16 years old.[70] This event centred on discussions about 'How can young people make good choices in an online world?', and 'What solutions can support young people in making good choices?'.

Case Study: Understanding the behavioural drivers of corporal punishment and co-creating an evidence-based solution with refugee teachers in Tanzania

From 2016-2018, BIT partnered with the International Rescue Committee (IRC) to develop a behaviourally-informed approach to reduce violence against children in and around schools in refugee camps in Tanzania. Most of our work took place in Nyarugusu refugee camp, the third largest refugee camp in the world and home to nearly 140,000 refugees from neighboring Burundi and the Democratic Republic of the Congo.

Given the complexity of the challenge, we started with a broad Explore effort to understand the types of violence that were prevalent, where they took place, who was involved, and what behavioural change would be needed to improve children's lives. This included two rounds of semi-structured interviews with teachers, caregivers, school administrators and children, as well as a review of the violence prevention literature, with a focus on violence in East African schools.

Through our Explore work, we narrowed down the focus of our project to violent discipline by teachers in schools. We learned that many teachers living in Nyarugusu thought harsh punishment was socially acceptable because they believed it prepared students for adulthood, taught them to respect their elders, and guided them to a better future. The difficulty of working in overcrowded classrooms with scarce resources and limited training also meant that teachers needed a quick, effective way of de-escalating misbehaviour. Many of them reverted to the same techniques they had experienced in their own childhood: 'the stick', a branch used to hit children, along with other physical forms of punishment. Teachers feared losing respect by failing when trying new strategies.

Our Explore work taught us that shifting teachers' attitudes towards the use of violence would be important to sustain long-term change. Traditionally, organisations around the world have deployed codes of conduct, as well as campaigns to raise awareness of current rules and children's rights. But our research suggested that these methods had not been very effective at preventing violence in Tanzania.

This diagnosis guided our evidence review towards strategies that had worked elsewhere to overcome the key drivers of violent teacher behaviour. For instance, we know that substituting habitual behaviour is easier than stopping it altogether, especially when teachers still faced the challenge of managing large classrooms every day. The Good Schools Toolkit, created by Raising Voices in Uganda, offered inspiration.[71] We introduced new concepts through edutainment,[72] in the form of stories, inspired by anecdotes shared by refugee teachers during interviews. These short narratives conveyed complex ideas and new perspectives in digestible, concrete terms.

We designed a rigorous trial to test this traditional strategy against strategies that focused on building empathy and making the harm salient. We found that building empathy for children's experience was the most effective way of getting teachers to start disagreeing with the use of corporal punishment.[73]

A subsequent independent evaluation found that the overall intervention was not effective in reducing levels of violence at 2 and 10 months after the intervention implementation. However, the process of developing the solutions gave our team many ideas to base future evaluation on and the results represent a critical addition to the small (but growing) body of evidence around violence prevention. [74]

Observation & Participation

When trying to understand a process or system, asking people about it may not be the best approach. Even when they intend to give you accurate information, it can be difficult for them. They may have forgotten, feel the need to tell you what they think you want to hear, or may simply not be aware of their actions..

One of the most illuminating approaches to Explore work can be to observe or participate in an action, service, process or system. This, combined with, for example, interviews with people to uncover their motivations, allows you to understand what they do in greater depth than through second-hand information. It also tells you about the process or system's context, and helps you identify differences between its design and implementation.

For example, observing a process might uncover efficient workarounds used by staff, and highlight opportunities for streamlining services. And participating in the process might reveal why something takes a long time to complete, or unpacks other 'hassle factors' highlighted during any interviews you have conducted. All these 'frictions', although seemingly small, can have a large impact on behaviour.[75] We have seen this time and again in government services, in everything from police cells to online services. Observing and/or participating is an excellent way to understand these frictions (Box 10 describes some of the behavioural evidence behind this).

Box 10. Understanding a system through the fundamental attribution error

The Fundamental Attribution Error is a wonderfully named (but controversial)[76] cognitive bias. It describes a phenomenon in which observers trying to explain someone's behaviour will often fail to be sensitive to the constraints of the situation or the environment, and will instead attribute the behaviour to a personal disposition.[77] However, the reverse is true when someone is explaining their own behaviour: they will emphasise the situation. Later work has argued that the type of outcome matters. If something positive happens, people are more likely to emphasise how critical their personal attributes were. If something negative happens, then it is not a reflection on their poor effort or ability, but the situation.[78]

Regardless of what happens, participating in a system forces you to understand the system from the viewpoint of the actors in the situation. If the research underpinning the fundamental attribution error is to be believed, being a participant, rather than an observer, will force you to pay more attention to the situational factors that affect people's behaviour than you would if you were just an observer. This process will lead to richer Explore findings.

Step 1. Choose a technique

In this section, we'll talk through two different ways in which you can experience a system. The first involves experiencing it vicariously, through **observation**, and the second involves experiencing it directly, through **participation**.

Observing involves seeing how end users experience a process or service, and how frontline staff deliver it. One way of doing this is 'shadowing', where you trail a person to understand their day-to-day routines and interactions. (With their consent, of course). We have done this in places ranging from job centres to courts and have 'shadowed' frontline staff such as court bailiffs and safety inspectors.

Participating involves going through a process or service to experience it as an end user. We have done this in ways ranging from applying for education scholarships and switching energy providers, to going through a mammogram, in order to understand what these activities required and felt like. This method is sometimes referred to as a 'service safari', but it is also known as 'mystery shopping', 'participant observation' or even 'immersion journalism'.

Which of these you choose should depend on your project. You may only be able to observe, rather than participate. For example, in a hospital, you can observe staff on a ward round, but you are unlikely to be able to participate as a patient. For this reason we have combined observation and participation into one tool. Alternatively, you may be able to do both at the same time. For example, if you are trying to better understand public transport congestion you can observe staff while using the services yourself.

Step 2. Plan your field visits

If a process is online or through a teleservice, you may be able to conduct your field visits from home or the office. During the Covid -19 pandemic, many processes moved online, which meant that it was also easier to conduct online fieldwork from home. However, you should be aware of the differences that might exist between your access to the internet and the people you who might be trying to help. For example, walkthrough videos on how to apply for homelessness support in 4K resolution may look great for someone with a reliable and high speed internet connection, but may pose a problem for people whose data plan only gives them 1GB a month.[79]

In many cases, Explore work occurs out in the field. Therefore, you need to identify the best locations to visit. You should start by looking at what you already know based on existing questions and data. This can help you choose sites that will give you a well-rounded picture. Think about experiencing:

- **Different parts of a system.** For example, if you are interested in social housing rental arrears, you may want to observe the sign-up process, (including any follow-up communications on how to pay rent and what happens if you do not) and late payment visits.
- **Sites with different operating models and outcomes.** Services often differ in how they implement a process and in the outcomes that they achieve across sites. It can be helpful to observe or participate in the same process at a range of them. Conducting observations at sites with low, medium, and high levels of social housing rent payments may give you very different insights and will stop you from over-generalising your findings.
- **New locations.** If this is a process or service that you are very familiar with, going to a location you have never been to before will help you experience it with fresh eyes.

In addition, you need to decide how you will capture what you have observed. An observation sheet or Framework[80] can help you do this consistently. With explicit permission, you can take photos, or record audio or video, which can help bring your findings to life in reports and presentations. (We also describe a case study on how you can use CCTV that is being recorded routinely to do this below). However, be aware that people may behave differently if they know their behaviour is being recorded.

Box 11. Record your findings ASAP

In his classic 1930s The War of the Ghosts study, Frederic Bartlett uncovered how powerfully we can misremember.[81] By asking people to listen to a story, and then getting them to reproduce it in the following months and years, he showed how over time we revise our 'memories' to fit our cultural and personal expectations.

Interestingly, Bartlett also found that people retold stories of similar length over time, but many of the original details were lost. The missing parts of the original story were filled with confabulations. Importantly, it was the details that did not conform to standard Western storytelling narratives or 'schema' that were lost. In the context of fieldwork, this means that our recollections of our observations may lose some of the inexplicable complexities and quirks of the environment we are investigating over time.

The tendency to simplify stories, along with others, such as the tendency to remember more about the start and end of an experience (the primary and recency effect),[82] and Elizabeth Loftus's work on reconstructive memory[83] (and its recent replications),[84] reinforce why you should record your findings as close to your research as possible. Doing this, for example, immediately after observing and participating will help you minimise these effects and lead to a more accurate recording of your findings.

A key component of your planning should involve working out whether you will explicitly notify the people you are observing (or the people involved in the process you are participating in), or remain anonymous. Whether and how you choose to notify the people you are observing will depend on the project, your role and the broader policy context (for example, if you are visiting a local site but your day job is based in head office, then there are different considerations than if you are from an external research organisation).

If people know that you are conducting some observations, then people may to change their behaviour (often referred to as the Hawthorne Effect)[85]. For example, while being observed, staff may deliver a best-practice service, instead of showing you how a service is delivered in practice. However, there might be some ethical or legal considerations if you choose not to tell people that they are being observed. These will depend on the jurisdiction that you are in and you should be aware of the specific ethical and legal norms of where you are doing your research.

To minimise your impact, try to be as unobtrusive as possible. In more everyday environments you should dress and speak more casually than you normally would in the office. You can also try to speak to staff when they feel less like they are being scrutinised, such as during breaks and informal meetings.

In addition, you can minimise the effect through repeat visits — as you build trust and familiarity, people will most likely revert to their natural behaviours, and you are more likely to get an accurate picture as you become 'part of the furniture'.

Box 12. Using a blended or devolved approach to notification

You can also take a blended or delegated approach to notifying people. For example, in a blended approach you might let the staff in an employment service know that you are observing how they help their clients, but avoid telling the clients. Similarly, you can take a devolved approach in which you train internal staff to code certain observations. We recently used this approach when working with a large Australian bank to identify where people were likely to breach Covid -19 physical distancing guidelines. We were not able to view CCTV footage from within the bank's headquarters. Instead, we developed a coding framework in which security staff were asked to look at the CCTV images every 900 frames (one image per 30 seconds) and code the proportion of people standing within 1.5 metres of each other. This preserved anonymity, but gave us a rich data source.

Whichever approach you choose, make sure that you and your partners are comfortable with the ethics of the approach, as discussed in Part 1 of this paper..

Step 3. Conduct your observations

Once you have planned your field visit or service safari, you are ready to go. Whether you are conducting observations or a service safari, there are two key things you need to do: notice and record. It is very easy to do one at the expense of the other. It is really easy to spend all of your time with your nose in your notebook furiously scribbling down your observations and miss some of the key parts of a process. Similarly, it is easy to forget your observations because you were so deeply immersed in your site visit. Finding the balance comes with experience and deliberate practice, and will also depend on how good your memory is.[86]

You can also use an observation guide to focus your attention throughout observations. These are best used when you have a narrow and clear focus. For example, in a project in which we were observing the use of handwashing stations in Bangladesh, we gave observers forms to simply count the number of times certain stations were used.[87] This type of form can also easily be built into an app, which means that observers just look like they are looking at their phones and are less obtrusive than if they were to be holding a clipboard.[88] This method is less appropriate for projects where the specific behaviours you are looking for are less defined — for example, when you are trying to identify which work health and safety risks affect gig economy workers you may not be aware of all of the potential risks so cannot develop a form to count them (see the case study at the end of this section).

Using video or audio recording can be helpful here. However, make sure that you have permission to do this. Also be aware of the fact that it is possible to record too much data. The case study at the end of this section describes a project in which members of our Australian team signed up to become food delivery workers. We used an audio field recorder to capture the process. These recordings were very informative, but they were also long. This meant that the team had to set aside a large amount of time listening back to the recordings.

Box 13. Pay attention to these five things

While participating or observing, watch out for:

1. **All the small steps in the process or service.** From your preparation, you should have an idea of how it works in theory, but this is your chance to find out how it works in reality.
2. **Emotions.** Pay attention to body language, facial expressions and tone of voice to understand how a process or service makes people feel. If you are participating, your own emotions can be very evealing too.
3. **The surroundings.** The physical space where a process or service takes place and how it is configured can have an effect on behaviour.
4. **Staff and end user interactions.** Their length, frequency and content can affect how personalised or user-friendly a process or service feels, which can alter how engaged they are with it.
5. **Existing communications.** Letters, forms and scripts that are related to the service or process are all potential intervention avenues.

Additional Tools

To augment participating or observing, you may wish to use these additional tools.

Demonstration or Usability Testing

Demonstrations or Usability Tests are designed to help you understand how people really act and to identify any barriers that stand in their way. In them, you ask people to show you the steps they take to do a task and then observe them doing it. Alternatively, you can also ask them to teach you how to do the task. For example, you might ask them to show you how they access a website on their own phones, or to give you a tour of an area. One way we have used this technique is in a project with the Australian Energy Regulator. In it, we asked participants to find the best energy offers on a comparison website. We observed them as they did this, taking note of barriers they faced and their comments along the way.[89] We have also used this technique to observe how small businesses were accessing the guidance provided by the Pensions Regulator in the UK.[90]

Empathy exercises

Participating is often the best way to get an idea of how a service or process feels for service staff and end users, but it is not always appropriate or practical. In these situations, empathy exercises can help you achieve the same goal. These are ways of simulating what an experience is like for others. For example, by:

- wearing gloves when opening medication to understand the experience of people with reduced dexterity;
- eating a meal while propped up in bed to understand the experience of hospital patients; or
- taking photos or videos of the lower shelves of supermarkets or on a walk to school to see what is at eye-level for children.

One thing to remember when using empathy exercises is to take care when drawing conclusions. For example, you might decide to use a wheelchair for a week to understand the barriers that regular wheelchair users face. While this exercise may provide some insights, if you do not use a wheelchair regularly, your lack of experience and skill might cause you to misidentify opportunities or misunderstand barriers. In this (hypothetical) example, there is a risk that this approach could backfire because trying out a wheelchair is obviously not the same as having to use one every day. In this situation, we'd seek the advice and input from regular wheelchair users to make sure that we're getting input from people with genuine lived experience.

Online observation

Just like traditional fieldwork methods, online observation means venturing into the field to gain a deeper understanding of the people we are studying. The 'field' here, however, is in the digital space. Online observation can involve tracing the pathway of a hashtag; participating in or observing an online community; or using an online space (for example, a forum we have created) to learn more about our participants in the non-virtual world. It is an emerging area of research that allows us to uncover new kinds of insights that can be triangulated with insights gathered from more traditional 'offline' research methods, but also opens up avenues for using tools to understand online behaviour more directly.

High fidelity prototyping

In the previous section, we described low-fidelity prototypes. As your solutions become more concrete, prototyping sessions are still useful to iron out kinks. In high fidelity prototyping sessions, you should give participants a version of your solution and ask them to use it as they would in the field.

We have found the 'think aloud' technique to be very useful in these sessions. This involves asking participants to say whatever comes into their mind as they interact with your prototype. This should include what they are looking at, thinking, doing, and feeling. The think aloud technique has been compared against other techniques and found to lead to more and deeper insights.[91] It can also be useful to pay attention to non-verbal cues in these sessions. For example, when we were running a prototyping session for an early version of an app that we had developed to encourage people to seek mental health treatments, we asked a small group of people to go through the signup process. While the users were focusing on the phone, we observed their faces. Furrowed brows and double takes were always the first indicator of a poorly executed user interface.[92]

Case Study: Reducing domestic violence reoffending through observations[93]

In New South Wales (NSW), Australia, 7,700 domestic violence (DV) Orders are breached every year. This causes immense harm to families and takes up a significant amount of court time and police resources. In partnership with the NSW Government's Behavioural Insights Unit (BIU) and the Department of Justice (DoJ), we conducted a project aimed at reducing reoffending by improving how defendants engage with the courts.

As DV rates are higher in rural and remote areas, but absolute numbers are higher in urban areas, our Exploration consisted of visits to nine courts – five urban, three rural, and one in a regional centre – to get a well-rounded picture. In addition, we spent several days in rural and remote NSW to get a better understanding of the cultural differences and unique challenges of delivering services in these areas.

Actively participating in the court process was not an option, but we were able to observe court processes from several different perspectives. For example, we watched defendants and victims interacting with court and police staff, and observed cases from inside the courtroom and specialist victim support rooms. We also observed how paperwork was done and collected blank copies of documents generated along the way.

When visiting our court locations, we tried to be as unobtrusive as possible, and to not impact on proceedings. As courts are busy, public places, this was relatively easy, with some exceptions. For example, specialist services like Victims' Support are generally gendered, so these were only accessed by female members of our team.

In addition to observing in courts, we conducted observations relating to other points in the process. We observed a men's behaviour change session, and visited police stations, victim's support services and various local community spaces that included messaging related to domestic violence (for example, any public health campaigns). Because of the sensitive nature of the environments we were observing in, we were not able to take photos, videos or voice recordings. We were, however, able to write on a structured observation sheet, and we typed up notes as soon as possible after each visit.

Our observations yielded many interesting insights. For example, we saw that in many courts, DV hearings are all held on the same day, and every defendant is told to arrive at 9:30am. This was leading to large numbers of (primarily) men waiting together for their case to be heard, arguably creating a normalising effect whereby the increased visibility of community members waiting for their court hearing meant that having an ADVO was seen as 'normal'.

Around the courts, we also observed imagery that may have compounded this normalisation. Anti-DV ads in the bathrooms of nearby pubs and clubs featured photos of beaten women, which may have normalised this behaviour in high-risk spaces. Our key insight, however, was that the 'legalese' used in court was likely preventing defendants and victims from understanding proceedings. The way in which the professionals in court spoke meant that many defendants did not understand what they could and could not do. This also meant that some of the victims that were present at the hearings were also not able to understand the protections afforded to them. In addition, the court forms were completely depersonalised, and used acronyms like 'PINOP' (person in need of protection) rather than the person's name.

Many court workers were completely unaware of the opacity of their terminology. Ironically, this phenomenon has many polysyllabic names and has been described by thinkers in a range of different disciplines. It was named 'trained incapacity' by economist Thorstein Veblen, 'occupational psychosis' by philosopher John Dewey, 'déformation professionelle' by sociologist Daniel Warnotte[94] or more recently 'the curse of knowledge' by Colin Camerer, George Loewenstein and Martin Weber.[95]

Building on these findings, in partnership with NSW BIU and DoJ, we redesigned the Apprehended Domestic Violence Orders (ADVO). We simplified and personalised the language, and added a message challenging the normalisation of violence. Our new plain-English versions also signposted support services and highlighted that this was an opportunity to change.

Following this project, the redesigned ADVOs were translated into 29 languages and rolled out across the state.[96] Our Explore findings are also being used as the basis for further promising interventions. For example, our findings informed the design of a successful RCT that used SMS reminders to significantly increase defendant attendance rates at court[97], and also led to the introduction of a brief court-based intervention with Aboriginal defendants which incorporates elements of behavioural commitment prompts and planning strategies.[98]

Case Study: Becoming a Food Delivery Worker

Food delivery workers in the platform-based gig economy risk significant psychological and physical harms on the job, particularly traffic accidents, verbal abuse and theft of delivery equipment.[99] While the number of incidents reported to the health and safety regulator in New South Wales (NSW) is low, there has been a rapid increase over the past three years, from one incident reported in 2017 to 19 in the first half of 2020 alone. Tragically, a further five food delivery workers died in traffic accidents on Australian roads between September and November 2020.

In 2020-21, we worked with the NSW Centre for Work Health and Safety (WHS) to gain a more nuanced understanding of food delivery workers' WHS knowledge, behaviours and concerns, and to develop a risk-reduction intervention that we could trial. We triangulated data from multiple methods and sources in our Explore work. A key part of this work was a service safari. This involved members of our team becoming food delivery workers to better understand the real-world decisions and risks they face.[100]

Three members of our team signed up to different food delivery platforms. They each went through onboarding, and worked at least one shift as a food delivery worker. During onboarding, the primary purpose was to identify what WHS information was provided by the food delivery platform, as well as any other instructions on how to undertake the food deliveries. When the team were completing their shifts, the primary purpose was to identify the WHS risks and any barriers and facilitators to mitigating those risks directly from the perspective of a food delivery worker.

The service safari helped us directly observe some of the risks that food delivery workers had reported in our earlier surveys. For example, only 5% of respondents reported 'never' rushing to deliver orders, and more than a third (35%) reported 'usually' or 'almost always' rushing to deliver orders. In our service safari, we saw many food delivery workers traveling very fast, including in pedestrian-only areas.

The service safari helped us understand some of the factors behind this self-reported rushing. Five key findings are shown below.

- **A lack of time:** For example, the time given to complete deliveries was sometimes less than Google Maps indicated was possible. This time allocation also did not give time for our team to lock up their bicycle or find a parking space.
- **A lack of feedback:** In addition, our team noted that they could not accurately gauge how their performance was being evaluated or see any consequences (either positive or negative) of their risky behaviour.
- **Using a mobile phone:** Using a mobile phone while riding or driving is a well-established road safety hazard. However, food delivery workers had to use their phones constantly while working, distracting them from monitoring the surrounding road and traffic conditions.
- **Using headphones:** The team members on bicycles needed to wear earphones to hear delivery notifications and route navigation instructions. This meant that they were distracted via visual and auditory channels.
- **Weather:** One service safari participant completed his shift during heavy rain, which necessitated regularly wiping off the phone screen with his sleeve.

Our service safari highlighted that food delivery workers make hundreds of small decisions over the course of a shift; distilling these decisions down to a series of interview questions would not have revealed the complexity, pace and interdependency of these decisions. In the context of WHS, even a small decision can have an outsized impact. A food delivery worker may be hit by a car as a result of an initial decision to cycle faster than usual to make a delivery within a short time frame. We used our newfound knowledge to co-design an intervention to reduce WHS risks on the road. This intervention was closely aligned with food delivery workers' actual needs and could be implemented at critical time points within the food delivery platforms' choice architecture.

Surveys

To understand and compare what different people think and why they undertake certain behaviours, we often turn to surveys. Surveys allow you to quantify and compare what people say, to get a sense of broad trends, possible causes and solutions, and the magnitude of a problem.

If you have conducted any research, you may have used surveys before. They are now relatively cheap and easy to run online, and governments around the world use them extensively. One of the benefits of surveys is that you can access the views of a large number of people within a few days. In many cases, you can achieve a representative sample. This means that you can make inferences about the wider population based on your sample. For example, if 20% of your respondents state that they have experienced sexual harassment at work, then you might reasonably conclude that 20% of workers outside of this sample have also experienced sexual harrassment.

However, if your sample is not representative, then you can make incorrect inferences. For example, if women are more likely to complete your survey and women are more likely to experience sexual harassment than other genders, then your inferences will be biased. This can be exacerbated by low response rates.[101] Despite this, surveys are still an important part of the Explore toolkit. Simple checks and balances can avoid issues related to representativeness and you should always triangulate what you learn with results from other tools.

Step 1. Identify your sample

First, you need to identify who you will survey. This group is known as your sample. Broadly speaking, the larger your sample, the better; but the right size for your project will depend on many factors, such as how many responses you expect, and how many groups you want to compare.

You will also need to consider your sample's representativeness. This is how much it mirrors the target population in factors including their demography, beliefs and behaviours. Importantly, sample size is no guarantee of representativeness. Depending on how respondents are recruited, large samples can still be biased.

To achieve a representative sample, there are a number of sampling methods you can use. Using these methods will also enable you to draw statistical inferences from your sample's responses. They include:

- Simple random sampling — choosing individuals randomly and with equal probability from the population you are interested in.
- Quota sampling — dividing a population into subgroups, and then sampling until these subgroups are filled.
- Survey weighting — applying a heavier weighting to survey results collected from underrepresented groups (useful if stratified sampling is not possible).

Alternatively, if you do not need a representative sample (you may only be interested in a specific segment of the population), there are other sampling methods you can use. These include Purposive Sampling where you target a specific group, or Convenience Sampling where you choose easy to access respondents. These methods are covered in Part 1.

Box 14. Asking about sensitive or socially desirable topics

Like interview participants, survey respondents can under-report their behaviour when asked questions about sensitive topics, such as criminal or sexual behaviour. Likewise, they tend to over-report when asked about socially desirable behaviours, such as using seat belts or paying taxes. To reduce this bias, you can:

- **Place sensitive questions around the middle of the survey.** Starting with difficult or very personal questions can put people off, and ending with them can leave the respondents feeling offended.
- **Divide sensitive questions into several parts that are individually less sensitive.** This makes the questions easy for respondents to answer, and increases the likelihood that you will get more accurate answers.
- **Make people feel comfortable telling the truth.** If people are likely to distort their answer to provide a more socially desirable response, you can put them at ease by saying that all responses are acceptable.
- **Ensure anonymity.** To encourage more accurate responses, you can assure respondents that it will not be possible to match their identity to their answers. It is important to note, however, that this may not always be effective. Work by Leslie John, Alessandro Acquisti and George Loewenstein, has shown that when privacy assurances are prominent, students are less likely to share personal information.[102] This finding suggests the need for a balance between assuring confidentiality and not prompting respondents to be unduly concerned about their privacy.

There are various complex methods to elicit truthful responses to sensitive questions. These include the unmatched count technique,[103] the bogus pipeline[104] or the randomised response technique.[105] They rely on a number of assumptions that we will not cover here, but might be worth exploring.

Step 2. Select your questions

Developing a good survey is a real skill. How you frame the issues, how many questions you choose to ask and the order you ask them in will affect how people respond. With structured surveys, you also usually only get one shot, and cannot clarify or ask follow-ups.

When you are developing a survey you should always refer back to your broader Explore questions and think about what you are actually trying to measure. In most of our Explore projects, our survey questions relate to very specific topics that are only applicable to the policy or service that we are studying (i.e. which parts of a service people find more useful). In other cases, we might be looking to measure deeper 'constructs'. Broadly, the term construct relates to a theoretical attribute that you are trying to measure.[106] Common constructs that you might want to measure include: life satisfaction, job readiness, depressive symptoms, personality, intelligence or burnout.

If you are trying to measure a broader construct, one useful tip is to use established or validated ones from research organisations, instead of writing your own questions. The UK Data service, for example, provides access to a large collection of surveys with topics ranging from volunteering to crime.[107] Similarly, there are many question banks for understanding people's personality types[108] and childrens' character skills.[109] Using one of their surveys also provides you with a ready-made comparison group. For example, our Behavioural Insights Team internal staff survey uses items from the Civil Service People Survey.

Box 15. What do we mean by a validated and reliable survey?

Whenever you are using a survey to measure something, you should be sure that you are actually measuring the construct that you hope to capture. Checking the validity of a survey can help with this. You can either do this yourself (or with the help of a psychometrician), or simply use a validated survey.

There are various different types of validity which are worth assessing. We list four types below. The first is **content validity**, which indicates the extent to which a survey measure or scale represents all facets of a construct based on someone's judgement.[110] For example, if you are trying to measure happiness, does your survey ask people about pleasure (hedonia) or their sense of fulfilment (eudaimonia)? The second type of validity is **convergent validity**. A survey measure or scale has this if someone's responses to it are similar to a scale that seeks to measure a similar construct. For example, if you have created a short survey to measure wellbeing, you would generally expect people's responses to it to be correlated with their responses on a similar measure developed by someone else.

A third type of validity is **discriminant validity**, which indicates the extent to which someone's responses to a survey are not correlated with something is conceptually very different or something that is conceptually the opposite of what you are trying to measure. "For example, if you are using a scale to measure happiness then most people who score highly on it should also have low scores on a survey that measures unhappiness. The fourth is **predictive validity**. This is the extent to which your measure predicts something in the future. For example, if you are trying to measure job-readiness, then the people who score highly on the measure should be more likely to find a job within 6 months than those who had low scores.

You may also want to look at the reliability of the measure. A reliable measure will generally give consistent results (even if it is consistently measuring the wrong thing). Reliability can refer to consistency over time.[111] For example, you would not expect someone's personality to change dramatically over a week. Or it can refer to consistency within the survey itself. For example, if you have a survey in which you are measuring anxiety with nine different questions, you would expect all of these to be correlated with each other.

If you do choose to use a validated survey, we will give you a word of warning. The psychometricians that develop surveys are often quite verbose, so you can end up with a very long survey. You should always consider what this might do to your response rate.

If you decide to create your own survey, an important choice to make is whether to ask open or closed questions. Open questions, which respondents answer with their own words, allow for in-depth responses, but we tend not to use them as they take longer to complete and are not quantifiable. Closed questions, which ask respondents to choose from pre-written answers, allow you to easily compare responses, but can inadvertently exclude answers not on the list. If pre-coded answers are to be used, it is important to run pilot surveys to establish what these are. The ideal question type, or combination of types, for your project will depend on many factors, such as the purpose of your survey, the complexity of your topic, and how much time you have to analyse responses. Regardless of what you decide, the checklist below will help you write good questions.

Box 16. Writing survey questions

Good survey questions are easily understood, are interpreted in the way you intended, and do not bias the respondent's answer. To make sure your questions meet these criteria:

- **Use simple language.** Using common words like 'drink', instead of words like 'beverage', makes it easier for respondents to understand your questions. Similarly avoid long or complicated wording, and be especially careful to avoid double-barrelled questions.
- **Use neutral language.** Generally speaking, avoid emotive or judgemental language. It is better to ask respondents to, 'Rate how informative this nutritional label is on a scale of 1 to 10', instead of asking, 'How confusing do you find this nutritional label?'.
- **Avoid homographs.** Words that are spelled the same, but have different meanings, can sometimes lead to confusion. Examples include words like tear, lie, and minute.
- **Be specific.** Specific questions are more likely to be interpreted the same way by all respondents. It is better to ask, 'How many times did you leave the house this week?' rather than 'How often did you not adhere to self-isolation rules this week?', as people may interpret what constitutes non-adherence differently.
- **Use a relevant reference period.** Using one relevant to your question will help respondents answer accurately. For example, it is better to ask how many soft drinks they drink each week, rather than each year. Try to stick to the same reference period, or response scales, across questions.
- **Allow for disagreement:** People often agree just to avoid seeming disagreeable. It can sometimes be better to use an attitude scale (strongly agree to strongly disagree) rather than asking respondents if they agree with a statement.
- **Work out if you need a neutral response option.** 'Fence-sitters' may skew your results if they are forced to choose between opposites. You can avoid this by including a specific option for them. In some cases, you might think that it is important to push fence-sitters to pick a side; in that case, you might want to avoid the neutral option.
- **Provide a 'Do not know' option:** 'Floaters' are respondents who offer an answer, but would choose 'Do not know' if they could. In a typical survey sample there can be many respondents like this, so including this answer for them is important.

Lastly, questions can be tested before going live, especially if new or innovative questions (and response categories) are being used. This allows you to check that respondents can clearly and rapidly understand what is being asked, do not find the questions or answers offensive, and that the question flow feels comfortable and not confusing.

Step 3. Choose your format

Your final task is to choose how you will deliver your survey to respondents. The format can have a big impact on the type of information that you collect, your sample and the way in which you can interpret their responses. Even different survey companies, asking the same question, can get different answers due to differences in the channel or samples used. Your main choices, and their biggest pros and cons, are:

Table 4. Pros and cons of different survey formats

Format	Pros	Cons
Postal	• There is some evidence that they obtain a higher response rate than online surveys.[112]	• Respondents cannot ask for clarification and need motivation to return the survey. • May under-represent younger people or other populations with less fixed addresses. • It does not capture metadata like the time it took to complete the survey. • You will need to convert the data into electronic form to analyse.
Phone	• It can obtain responses from people who are less familiar with online data collection methods. • Respondents can ask for clarification and interviewers can ask probing questions.	• Are resource intensive. • Fixed line phone surveys are no longer representative, and mobile numbers can be hard to obtain. • Response rates for phone surveys are dropping.[113]
Face-to-face	• Respondents can ask for clarification. • Interviewers can probe and use additional resources.	• Very resource intensive. • Subject to interviewer bias. • Can remove the perceived anonymity of respondents.
Online	• Are low cost, with wide reach and fast turnaround times. • Offer greater control with the option to randomise or set mandatory questions. • Will capture meta-data like time the data was collected and how long it took.	• Respondents cannot ask for clarification. • Can be hard to ensure the correct sample is reached.
Text/SMS	• Are low cost, with wide reach and fast turnaround times. • Forces you to ask only the most important questions.	• Restricted to very short surveys. • Potential biases in response rates, depending on mobile phone use.
Digital feedback terminal	• Can be targeted at specific moments/point of service.	• Are largely restricted to single question ratings.

Box 17. Increasing response rates

One of the biggest challenges with surveys is getting enough people to complete them. To maximise your response rate:

- **Keep your survey short.**[114] One study found that reducing survey length from 20 minutes to 10 minutes led to an increase in completion rates from 57% to 68%.[115]
- **Make it easy to complete:** Only ask what you need to know, use simple language, ask multiple choice rather than open-ended questions, auto-populate fields where possible, and break long lists down into shorter, discrete tasks.[116]
- **Think about the survey's user experience:** Start the survey with easier questions. Leave difficult or demanding items to the end.[117]
- **Make it mobile-friendly:** Ensure that your survey is optimised for mobile devices and avoid formats that are difficult to use on small touch screens.[118]
- **Be clear in your ask.** Let people know how long it takes to complete your survey (and be honest about it) and make them aware of the deadline for completion.
- **Grab their attention.** For online surveys, including a preview of the first question in the invitation increases response rates.[119] For paper-based surveys, emphasise that there is a real person behind the message. In past studies, personalised, hand-written Post-It® notes stuck to the front page of the survey resulted in more surveys returned, surveys returned faster, and more detailed responses.[120]
- **Use incentives.** The evidence is mixed, especially for online surveys, but offering financial or non-financial rewards can help to increase responses, especially if these are pre-paid.[121] Studies have also shown that lottery incentives with small chances of winning a large prize can be effective.[122]
- **Make it timely.** Contacting respondents (for example,by SMS) ahead of sending or emailing them a questionnaire has been shown to boost response rates[123], as can sending up to three reminders[124]. With online surveys, the time of day can also make a difference. SMS and emails are more likely to be read and responded to if they are sent before or after work, or during lunch.[125]

Alternative methods

In addition to the more traditional survey methods covered earlier, there are alternative methods you can use to understand what people think and do.

Online citizen panels

Sometimes used by governments and councils, online citizen panels allow people to provide policy advice on issues to do with their local communities through surveys, discussions, and other means. These are increasingly occurring online, partly due to the Covid - 19 pandemic, but also for cost reasons.[126] These panels do not need to be solely online. BIT has run numerous panels which have combined online and offline panels.

Studies have shown such panels to be effective in engaging communities that were previously labelled 'hard-to-reach' by government. The World Health Organisation have recently recommended their use to encourage social participation in health policy development.[127] A pilot involving three South Australian metropolitan councils found that they engaged people who had never been involved with their council before.[128] This highlights that the terminology of 'hard-to-reach' populations is flawed. It is not that the people are hard to reach, but that the tools being used to reach them are inadequate.

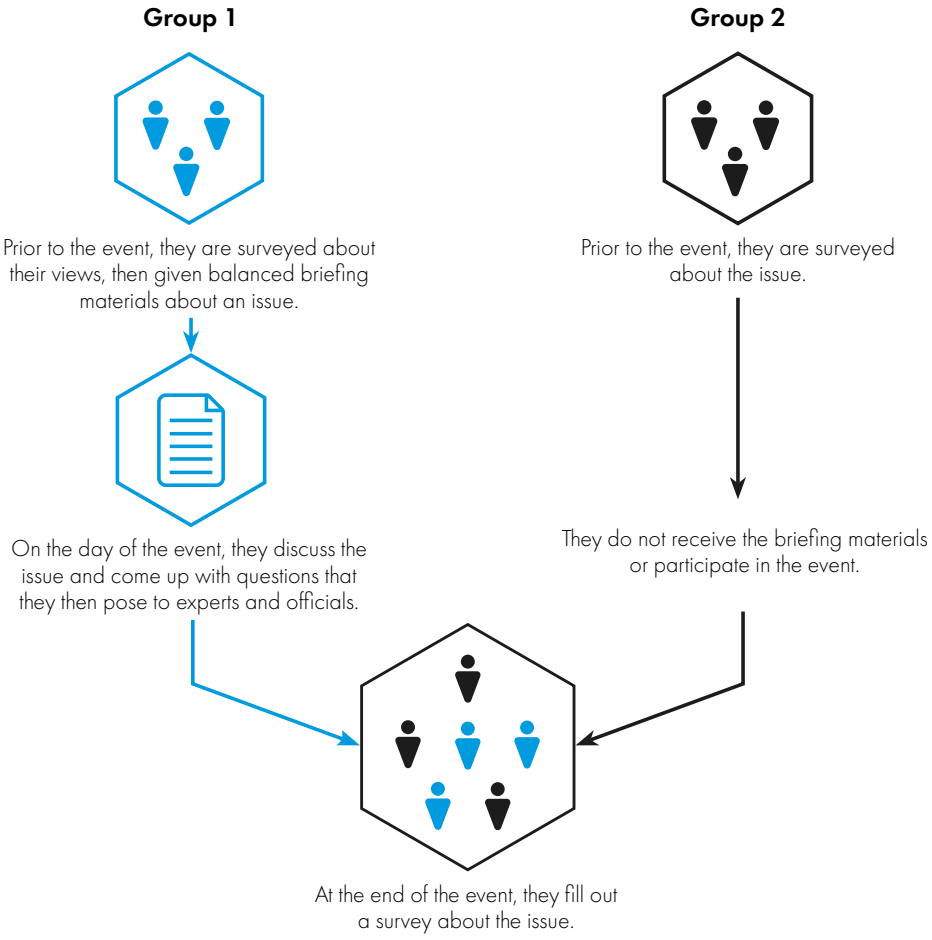
One issue to watch out for, in common with most panel surveys, is that repeated questioning of the sample can make them less representative of the general population over time. Being asked about council services on a regular basis, for example, can lead the sample to become more knowledgeable about and more engaged in them. This is sometimes corrected by gradually refreshing the panel, i.e., adding new respondents and gradually retiring long-standing ones. Having increasingly engaged, and therefore less representative samples, may not be always a bad thing. Their newfound engagement with the issues that you are Exploring might make them effective advocates for change or allow them to see the issue in a different light, giving you deeper insights. You just should not consider them representative of the wider community.

Deliberative polling

Deliberative polling combines techniques from public opinion research and public deliberation. It is designed to assess what public opinion on a particular issue might look like if citizens were given a chance to become more informed.

The term 'deliberative poll' is associated with a particular method and has some subtle but important differences from other deliberative mechanisms.[129] Just like a conventional survey or 'poll', a deliberative poll typically involves several hundred people and careful efforts to make sure the sample is representative of the wider population. Broadly speaking, it involves two representative groups of people chosen from the general public who go through two different processes:

Figure 2: The different processes that the groups in a deliberative poll go through



After the event, all the survey results are then analysed to see if the additional information given to Group 1 affected their opinion on the issue. Alternatively, in some versions, only 'Group 1' is used, and the change in their average opinions is the measure of interest, i.e., suggesting how views of a wider population would change if exposed to similar information and debate. The 'Group 1 only' process was used by the UK Department of Work and Pensions to assess public opinion on proposed Pensions Reform. Their research found that the deliberation led to 80% of the panellists holding 'fundamentally different' views at the end of the day to some of the proposals.[130]

Online experimentation

Another option is to combine survey questions with experimentation using online platforms like our own Predictiv. Surveys themselves are great at finding out what people's views are on given topics, but in some respects that's all they offer: a reflective response to a given question. But it is possible to sharpen this by combining surveys with experiments that can elicit preferences from competing options, present options randomly to choose from, or randomise people into different conditions and track answers following exposure to a given stimulus. We have spent time conducting numerous online experiments via our Predictiv platform in order to develop policy options, and in doing so worked closely with policy-makers and understand what is needed to do this well at speed.[131] Online experiments will never be able to replace testing a new policy in 'the field' (unless the 'field' is online, of course!). However, they have three main advantages that explain why they are increasingly being used. They are:

1. **Fast.** Once an experiment has been designed, results can be collected in days rather than weeks or months.
2. **Easy to implement.** It can sometimes be difficult to get a partner to agree to run a field experiment, especially if they stand to lose if the trial has a negative result. Online experiments help to circumvent these issues.
3. **Cost effective.** Although there are still costs associated with online experiments (not least the incentives provided to participants), they can be significantly cheaper than alternative experimental designs.

We have now run over 200 Predictiv tests with over quarter of a million participants across the globe. These tests have answered lots of different questions that it would have been hard to answer as quickly or cost-effectively using other methodologies. We have shown:

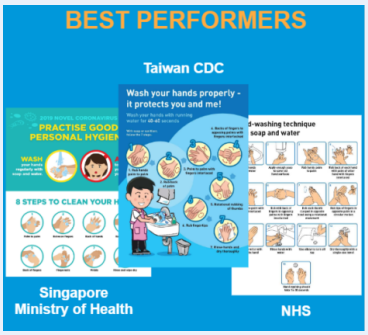
- that introducing sliders and changing the default minimum payments on credit card statements can help increase the monthly amount that people pay;[132]
- which messages are more likely to get people to self-isolate at the start of the Covid -19 pandemic;[133]
- that including an educational flyer to address common mistakes can help to reduce errors when filling out an Indonesian tax form;[134] and lastly,
- that making the Bank of England's interest and inflation rate announcements more relatable to everyday decisions can significantly improve how UK citizens understand them.[135]

Case Study: Using Predictiv to understand the impact of Covid -19 communications

At the start of the Covid -19 pandemic, the task of educating the public about what they needed to do to stop the spread was foisted upon many governments. The science on what behaviours were most effective in stopping the spread was fast evolving, so any communications that aimed to educate the public needed to be developed and adapted in near real-time. This gave little time for checking whether or not the public health communications used to educate the public were actually effective.

While BIT would traditionally advocate for an evaluation method that measured actual behaviour, we knew that this was not possible. However, a traditional market research methodology that relied on focus groups would not give an indication of which messages would be most effective. Thankfully, our Predictiv platform presented itself as a way to quickly identify whether the messages being developed by governments were likely to encourage people to wash their hands, self-isolate and stay at home.

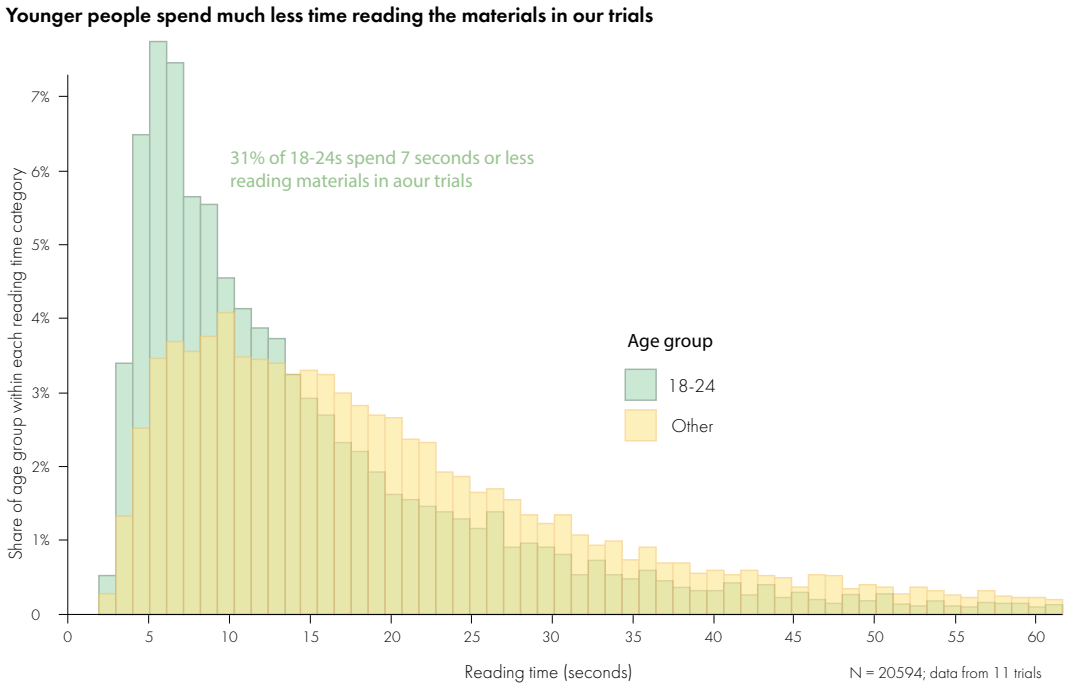
Throughout the pandemic, we worked with governments across the globe to test which messages increased the likelihood that people could recall the main messages in a public health communication, and whether it also increased their intention to undertake the relevant behaviours. Some of the messages we tested can be seen below. In most cases this involved recruiting a large sample of people to our survey and then randomly allocating them to see a selection of different messages, before then asking them a series of questions about the message.



As the pandemic progressed, our approach became more efficient and more sophisticated. We were able to test messages within 24 hours and combined the approach with virtual focus groups (often in different languages). These focus groups allowed us to generate quick insights on whether certain messages might backfire and whether or not some messages might not be interpreted the way people in government assumed they would.

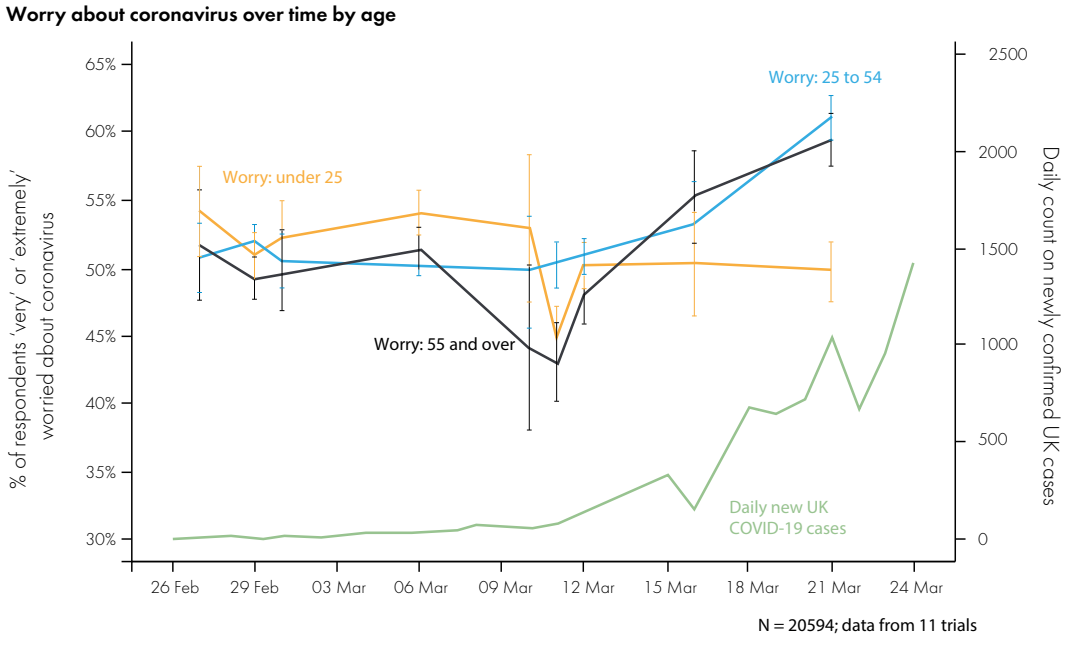
These online experiments also allowed us to track trends over time and conduct meta-analyses which identified trends across the globe. For example, we were able to identify the fact that younger people were generally less likely to recall the main messages from Covid -19 communications. They also spent less time looking at the materials we showed them. They also spent less time looking at the materials we showed them (see Figure 3). We were also able to see how people's risk perceptions changed as the virus spread across countries (see Figure 4).

Figure 3. Graph showing how young people differed from other age groups in the length of time they spent looking at Covid -19 posters



This approach has now been widely scaled across the globe. At the time of writing, BIT has run over 80 experiments on Covid -19 communications and has inspired a large number of other organisations and governments to run similar tests.

Figure 4. Graph showing how different age groups worried about Covid -19 over time alongside the number of daily cases seen in the UK



Administrative data analysis

In complex policy areas with rich data sources, where the level of detail and quantity of information can be enormous, data analysis can be incredibly powerful. Data analysis can reveal patterns and relationships that warrant further investigation; help you pinpoint where people drop out of a process, where or when you should intervene; and show you which groups to target to achieve the most impact.

It is also often best used in conjunction with other Explore tools. As mentioned, behavioural biases such as the social desirability bias can lead people to over-report good behaviour and under-report bad in surveys and interviews. Data analysis can help you identify when this is happening and give context about why. Like any form of research, however, it is not without pitfalls. Data analysis programs confidently churn out numbers, but it is up to you to ensure that they are built on reliable data, and that your interpretations of the patterns are well judged.

As a result, where possible, we recommend enlisting a data analyst. They can help you get the most out of your data, by identifying what it can and cannot tell you and calculating how certain you can be of your findings. With the rest of this chapter, we will outline the key tasks they can help you with, as well as the overall process from gathering data to finding insights.

Box 18. Data analyst as Explorer

The psychologist Raymond Cattell was one of the early pioneers of large scale data analysis. In particular, he used a form of data reduction called factor analysis to model the underlying patterns across all the data he could lay his hands on.

A metaphor Cattell was said to have used to explain his models to non-statisticians is particularly relevant to Exploring. He said to imagine you are a hunter in the jungle on the lookout for a leopard. By their very nature, leopards are hard to spot. As they slink through the undergrowth their camouflage helps them fade into the background, and most of the time they will be partially obscured by the brush and trees. At best, all you may see is the movement of some spots. How, then, will you know if you have seen a leopard?

Cattell argued that the key is to see that the spots move together. It is this pattern of linked movement that betrays that the spots are part of a whole and enable the hunter to 'see' the beast.

From the techniques that Cattell pioneered, to the cutting-edge machine learning of today, this is not a bad metaphor to have in mind as you hunt for insights in your data and assemble patterns from them. It is also a timely reminder to beware that the patterns you see may be at least partly shaped by what you think you are looking for.

Step 1: Gather relevant data

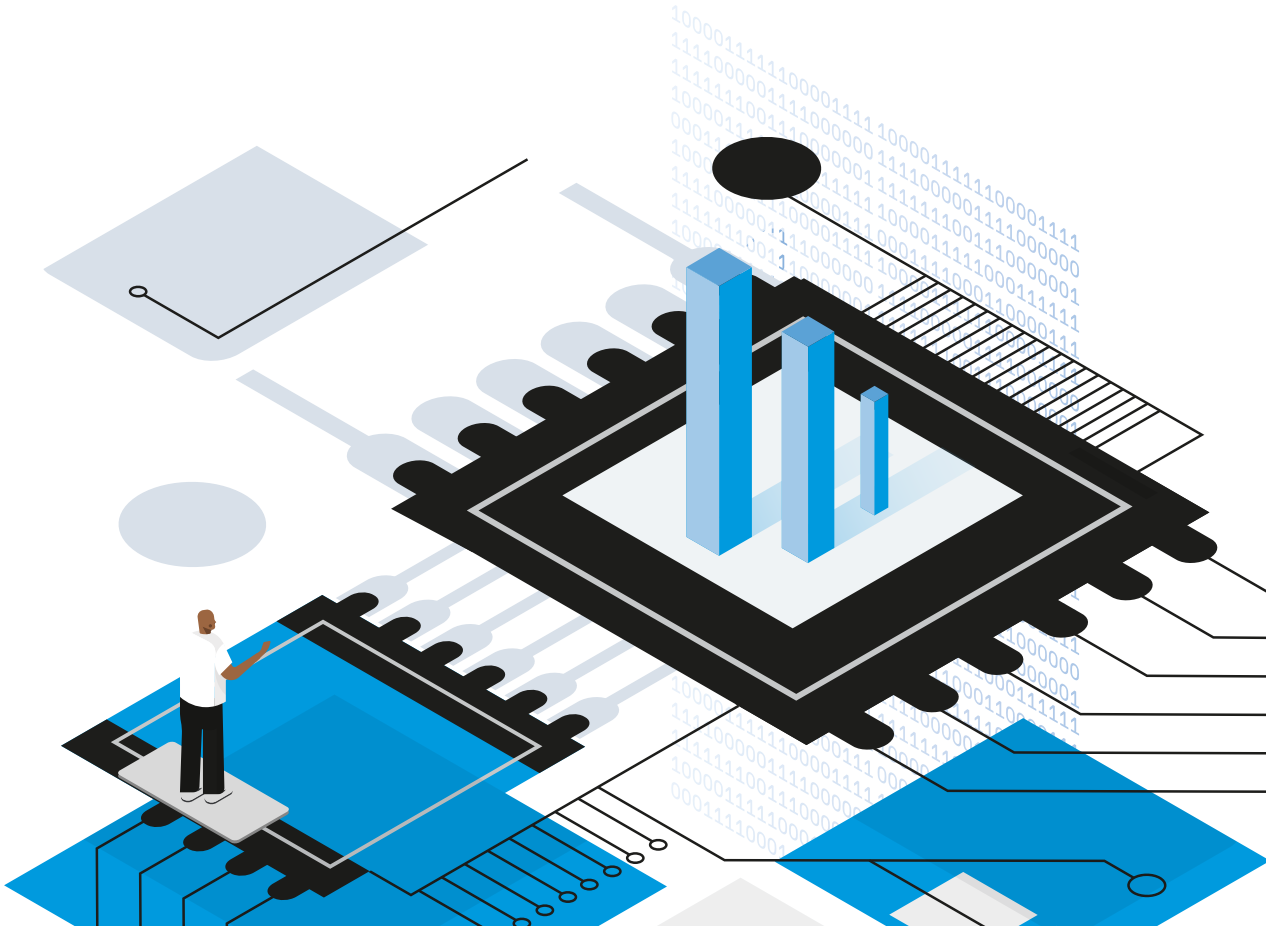
Your first task is to pull together all the relevant data that you can, from sources you have direct access to, and from colleagues and other sources you discovered while Exploring. This will allow you to link it together and uncover new insights.

Make sure to set time aside to do this analysis. You will need to work through the legal and ethical considerations inherent in working with data on real people. You should pay special attention to how data will be used and linked back to individuals. Removing a person's name also does not guarantee anonymity, so you will need good data security, encryption, and governance arrangements from the outset. Using synthetic data may also help overcome some of the ethical and legal issues of working with administrative data.[136]

Finally, when gathering data, try to identify someone with an in-depth understanding of the data being collected. This may be the owner of the data, or a data controller or analyst. You should then ask them to talk you through the dataset. This person may be able to tell you how each variable was recorded, to help you determine its quality and trustworthiness, and may be able to give you a Data Dictionary that lists the variables in the dataset and describes their format.

When working with administrative data it is good practice to observe it being collected (as far as this is possible) and asking the person to talk you through the process of collecting it. Speaking to a number of people who are responsible for collecting data will allow you to unpack the range of ways in which that same data is interpreted at the frontline. For example, if you are speaking to employment staff about how they enumerate the jobsearch activities that their clients are undertaking, you might learn that one staff member counts the process of downloading a job application, completing a cover letter and submitting the application as a single activity, whereas another might count the process as multiple activities.

Ideally, you should also ask the 'upstream' users of that same data across the organisation about how they interpret it. For example, how does the manager of the aforementioned jobcentre staff view the relative differences in job search activities in their caseload? Does the manager attribute it to differences in the jobseekers' motivations? It is worth undertaking this process at a few sites or with a few people as this will allow you to capture any variation in the way that different people capture data. Box 19 shows that it is not just people that vary in how they capture data, but machines too!



Box 19. Understanding how data is coded can save some embarrassing errors

Heating controls in UK households are often outdated and difficult to use. This led to the creation of smart thermostats that promised to reduce gas usage, lower carbon emissions and save people money. However, the effectiveness of these thermostats was not clear. So, BIT was commissioned to evaluate their effectiveness.

In the first step, we looked at meter data from energy suppliers. We discovered that this data was inconsistent. Because meters are denominated differently (some in cubic meters, some in cubic feet) and provide readings at different intervals (some daily, some erratically), and because they 'turn over' to '0' after their counter reaches its limit, we needed to piece together actual gas consumption from disparate parts.

When analysing the data, we found that some households in our sample seemed to be using too much energy — about ten times as much as an average house. Either everyone in the sample lived in a huge mansion with the heating on all day, or something else was going on.

A quick inspection of the data showed what was happening – where a meter reading had been missed, or recorded and subsequently corrected, it appeared in our data as 99999, and this was leading to the high energy usage levels we were seeing. This was confirmed when we spoke to our expert advisors. Not understanding how missing data is coded is a classic issue, but one that can be easily overlooked if you are in a rush to analyse the data. For this reason, we recommend speaking to the people who are involved in coding and collecting the data you are analysing, asking for a data dictionary whenever possible, visualising data before analysing it and running sense checks on the data that you have (for example, are people's energy usage levels in line with the population average?).

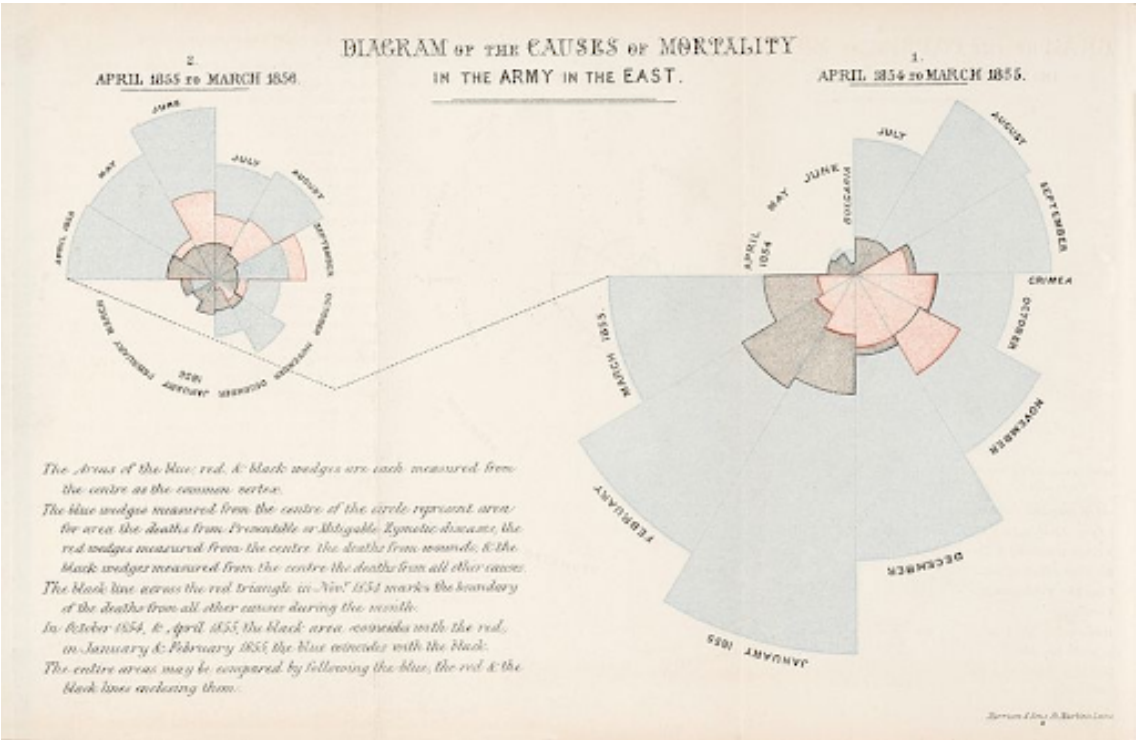
Figure 5. A reprint of John Snow's Map showing deaths from Cholera in Broad Street, Golden Square and neighbourhood

Step 2: Understand the data you have

Once you have identified what data you want and have gathered it, it is time to unpack what it can tell you, tidy it and transform it into something that you can analyse. This process can be broken down into these tasks:

- **Identify the key variables:** These are the outcomes you are interested in (for example, a person's employment status) and other variables they may potentially correlate with (for example, the region where they live). You should spend some time working out what your key variables of interest are and why. Your Explore questions will always be a good reference point.
- **Sense check the key variables:** Identifying issues with the data and fixing them early can save you a lot of time. This includes looking for missing values, impossible values (for example, fines paid before they are issued), or unexpectedly high numbers of particular values (for example, some survey designers code missing values as '999', which can throw your whole analysis out). This may seem dry, but it is incredibly important to check, or 'clean', your data before you start spending hours on the analysis.
- **Wrangle your data:** Usually when you receive any administrative data, it will be structured in a format that best suits its intended purpose (for example, seeing who has paid their tax on time). Administrative data is rarely designed in a way that allows you to analyse it immediately. Therefore, you should spend some time transforming the data into a format that will help you use it for research purposes. This process is often referred to as 'data wrangling'. Effective data wrangling is an art as much as a science, but there are good guidelines that can be used to make this process easier.[137]

Figure 6. Examples of the coxcomb or rose diagrams developed by Florence Nightingale



Step 3: Visualise your data

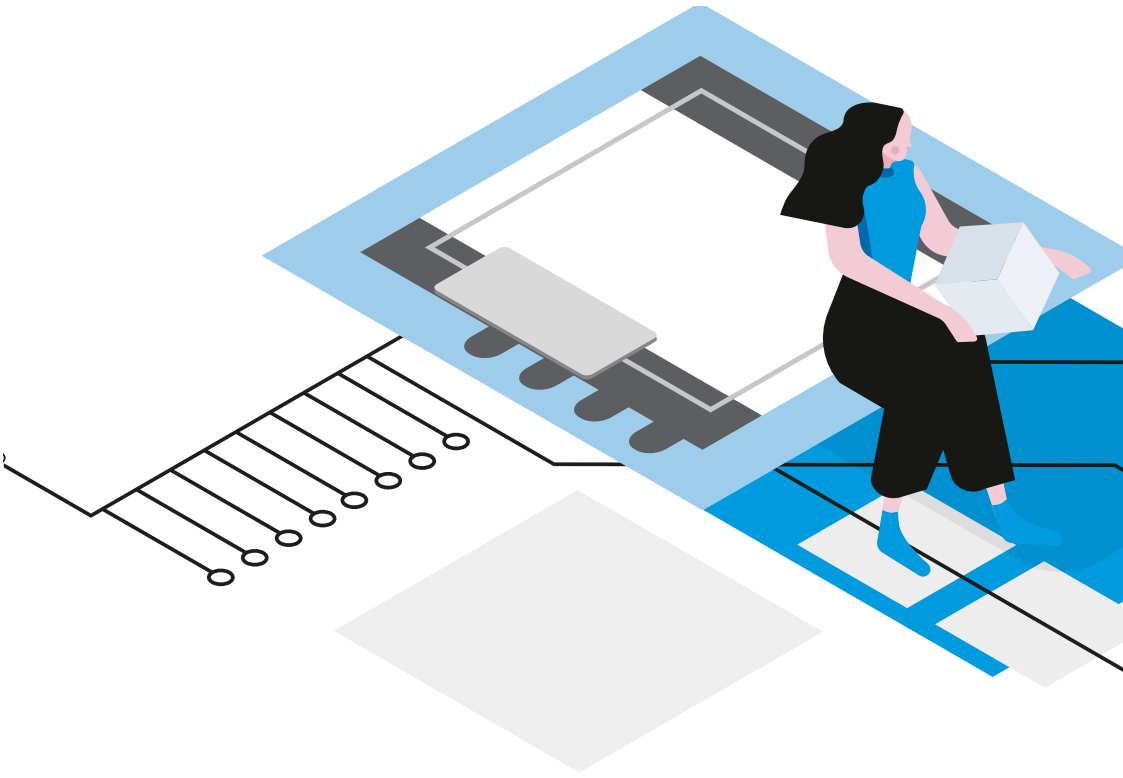
Once you have gathered your data and spent time trying to understand it on a theoretical level, it is really tempting to start running many different analyses to see what relationships might exist in the data. However, visualisations are usually a much better place to start. In 1962, John Tukey, one of the greatest statisticians of the 20th century, stated 'the simple graph has brought more information to the data analyst's mind than any other device'.[138] Nearly 60 years later, this statement should still be memorised by anyone who hopes to analyse any data.

Data visualisation can broadly be used in two ways. The first is discovery, and the second is communication. A great example of how data visualisation can be used for discovery is the famous cholera map developed by John Snow in which he reportedly plotted the locations of the homes of those who had died from the 1854 cholera outbreak in London. From the marks on his map, Snow could see that the deaths had all occurred near the pump on Broad Street and theorised that the cholera outbreak stemmed from the water source.[139] This map is reprinted as Fig 5.[140]

With these visualisations, you will be able to see how strong any relationships between your variables are, how data are distributed, and where there are outliers or clusters.[141] Importantly, many statistical models are based on assumptions about the shape of distributions and relationships, and can offer up misleading results if the assumptions are not met. Visualisations will help you understand the distribution of your data.

It is really important to consider how these visualisations sit alongside your other Explore findings and your hypotheses of what is happening. These visualisations can then be used to pinpoint opportunities to target your solution. It is the combination and triangulation of the data visualisations that gives the visualisations their power. This point has slowly vanished from the retellings of John Snow's cholera map. John Snow was only able to deduce that the Broad Street pump was at fault due to his carefully developed hypothesis that cholera was spread through water. Without this hypothesis, he might have blamed the butcher, the baker or a miasma emanating from the candlestick maker's workshop.

Data visualisation can also be used to communicate your findings. Florence Nightingale, as well as being widely seen as the founder of modern nursing, is increasingly being recognised as a trailblazer in using data visualisation to advocate for policy change.[142] Her use of rose diagrams or coxcomb diagrams, which have been reprinted as Fig 6, is seen as having paved the way for public health reform which saved and extended millions of lives.[143] This highlights the power of data visualisation to communicate your Explore findings.



Box 20. Principles of using data visualisation for effective communication

Good data visualisation can be hard to achieve, but there are numerous guidelines on how to achieve it. One of the most influential thinkers in this space is Edward Tufte.[144] In 1983, Tufte set out a number of principles for effective visualisation of quantitative data in his beautiful guide to visualisation,'The visual display of quantitative information'. These principles can be summarised as: "above all else show the data" and do so simply. Tufte's work implores analysts to focus on how complex information can be simplified so that the reader can gain insights.

The epilogue to Tufte's book invites the reader to greet his principles with skepticism and in the last 40 years, many have. Many of these critiques have focused on the fact that Tufte's emphasis on just showing the data does not necessarily improve comprehension or recall of the information shown.[145] This has led many researchers to look at how the human brain interprets visual information and how we can use this information to improve data visualisation techniques. This has shown that simplicity, on its own, does not necessarily lead to better communication.

The best method of communication will largely depend on what you are trying to communicate. If you are using the data visualisation to make a clear point, then a simple bar chart can be extremely effective. If you are using data visualisation as a method of exploration, then complexity may not be something to shy away from.

Step 4. Identify new insights

Once you understand the key features of the data and have used visualisation methods to unpack the data, the next step can be to apply statistical techniques to find new patterns and relationships; and uncover deeper insights.[146] We have purposefully avoided giving any detailed instructions on how to conduct statistical analyses in this section, as the topic has been covered very effectively in thousands of other guides.

If you are not a data analyst, this is where enlisting one's help can add a huge amount of value. Even if you outsource the analysis, however, you still have an important role to play – you need to guide them by asking the right questions.

The reason is, even with a small dataset, there are many possible correlations. For example, in a 25-question survey there are 300 associations to investigate. Not only will sifting through this many combinations take considerable time, there is also a risk that an unguided analysis will lead to spurious correlations.

The key point here is that before you ask a data analyst to look at your data, or before looking at it yourself, you need to have an idea of the hypotheses you want to test and the questions you would ask to test them. These can include the simple questions shown in Table 5.

Table 5. Types of questions statistical analyses can help you answer

General Question	Example question
What is the relationship between [Variable X], [Variable Y], and [Variable Z]?	How is school attainment related to parental background and quality of teaching?
Are there different kinds of [people/ behavioural patterns/organisations]?	Can we identify a certain number of skilled migrant archetypes?
Is there a significant difference in [Variable X] between [Group Y] and [Group Z]?	Which group found a job the fastest – the group that received the additional training, or the group that did not?

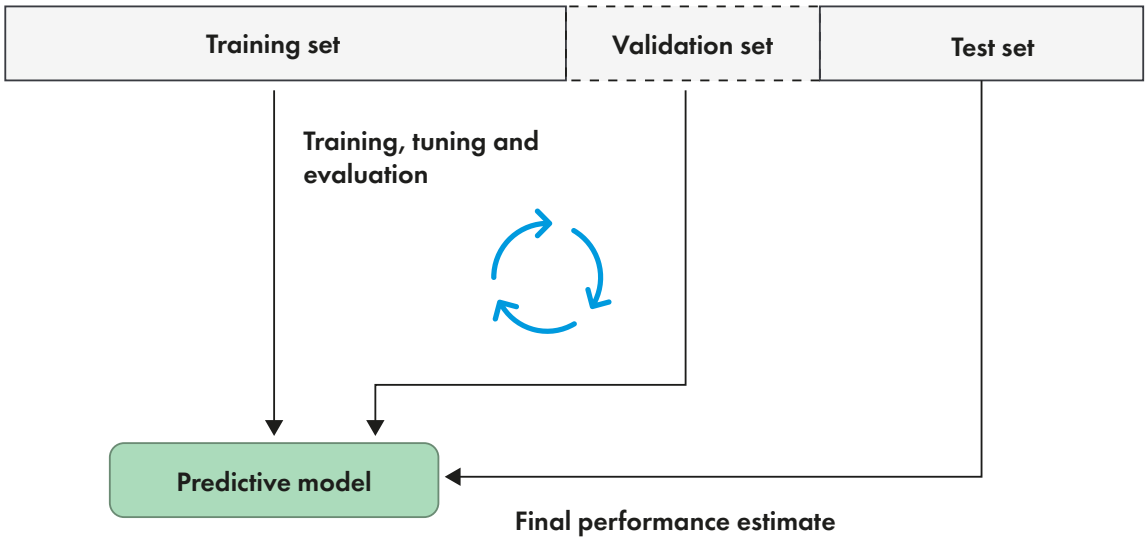
A key thing to bear in mind is the adage 'correlation does not equal causation'. Although it can be tempting to jump from the finding that 'A' is strongly associated with 'B' to the conclusion that 'A' caused 'B', it is important to remember that there are other possibilities. It could be that 'B' caused 'A' (for example, it may be that participants starting a new job may have led to them to do more training rather than better trained job seekers are more likely to get jobs), or that it was caused by an unrelated factor (for example,a large new employer recently set up shop nearby). Or it could just be a fluke. If we run dozens of correlations, some will come up 'positive' just by chance. The website (and subsequent book), 'Spurious Correlations', catalogues many of these flukes and the data manipulations required to achieve them, such as the finding that the number of non-commercial space launches correlates with the number of sociology doctorates awarded in the US.[147]

The opposite — causation does not imply correlation — is also true, but rarer. For example, it might be the case that a training program does help people find work, but as there are no vacancies in a location, you do not see an association between uptake of the program and higher employment.[148] If we think about this for a moment, it means that at a different time, an intervention that was once effective in the same location might no longer be effective because the external job market conditions have changed.

These points highlight how important it is to consider how you will interpret your data. You should always consider the risks of misinterpreting your findings. For example, if you are looking at a high stakes problem in which any unintended consequences have the potential to be severe, you should consider the limitations of your data (and therefore the certainty around your recommendations) alongside the potential costs and benefits of the policy response you are recommending. A poor interpretation of the data could have big implications for a large number of individuals, families and society as a whole. If you are making any policy recommendations, make sure that the strength of your recommendations is tempered by the strength of the evidence underpinning it.

For this reason, we have historically relied on evaluation methods like the Randomised Controlled Trial (RCT) to establish causality. We would recommend reading the short guide Test, Learn, Adapt if you want to understand how RCTs help establish the effectiveness of your solution.[149]

Figure 7. How data scientists assess the performance of predictive models



Box 21. A note on statistical significance

Anyone who has read a scientific paper will most likely have encountered the term 'statistical significance'. This term has a very specific technical definition, which can be described as: 'the probability of the results being this size or larger if the null hypothesis is true, where the null hypothesis is of no difference between the groups you are comparing.' We will not unpack this definition here, as many textbooks do this in more depth.

However, the meaning of this definition is often misinterpreted. It means that the results from a piece of data analysis generated by experimentation is unlikely to have occurred by chance. If something is statistically significant, it does not mean that a result is meaningful. Similarly, if a statistical test does not give a significant result, it does not necessarily mean that an important effect is not present.

The concept of statistical significance is generally only useful when a statistical test or experiment has been designed to test a specific hypothesis. Modern computing has made the running of statistical tests almost instantaneous. But statistical testing needs thinking through, especially when working with large samples (as we often do in policy settings).

Numerous people have tried to devise methodologies to distinguish statistically significant results from meaningful results. This can be particularly important in a clinical context - a statistically significant reduction in your symptoms may not make a material difference in your life. Therefore, Neil Jacobson and Paula Truax devised the concept of 'clinical significance'.[150] They also developed a simple methodology for establishing whether or not a change in outcomes is meaningful. Sadly, the specific rules can only be used when you are using a psychometrically validated survey as your outcome measure. However, the paper covers some important ripostes to the unthinking way in which statistical significance is used.

Similarly, Jacob Cohen developed many of the methods of power analysis in response to what he described as: "the ritualization of null hypothesis significance testing ...to the point of meaninglessness and beyond". Cohen's writing is very accessible and documents his increasing frustration with the application of null hypothesis significance testing[151] over the decades. We would highly recommend reading his later work.

Additional tool: Data science and machine learning for prediction

Data science

There is much excitement about a new generation of statistical techniques that combine the brute power of modern computing with modern societies' ever-growing volumes of data. The term 'data science' is sometimes used interchangeably with 'machine learning'. Definitions of data science and machine learning vary,[152] so it is always important to check whether your definition of data science and machine learning is shared with the people you work with.

Machine learning

Here we will use the term machine learning to refer to the techniques that are used, and data science as the broader field of study. At their core, these techniques use algorithms to build complex statistical models[153] to either make the best predictions about individuals' outcomes (prediction) or to identify the best way to classify people into different groups based on hidden patterns (clustering or segmentation).[154] Prediction often requires a predetermined outcome of interest (for example, tax payment), whereas classification does not.[155]

An example of machine learning techniques being used for prediction is a project we ran with the Sussex safer roads partnership in the UK. They asked us to explore the relationship between serious road injuries and driver, car and road characteristics. Using machine learning, we found that many of their beliefs about risk factors were not borne out. At the same time, we found other patterns, such as an elevated risk from drivers in their 50's, who had perhaps become overconfident and error-prone.[156]

A big risk for machine learning is that by trying out millions of predictive models, the machine might come up with an excellent 'fit' that is a total fluke. This is called 'over-fitting'. To overcome it, analysts can split their data into three subsets. They 'train' the model on the first training set. They tune it with a second 'validation' set to see how well it does.[157] This extra data then helps the algorithm learn an appropriate level of complexity. Once the algorithm has been updated, they use the final test data, which they held back, to test it, helping them determine the model's trustworthiness. Analysts will usually then state how well the model 'predicted' something about the test set. It is important to note that in this instance, the analysts are predicting events that happened in the past. This process is shown in Figure 7.

Machine learning can be an extraordinarily powerful tool. And like all powerful tools, it has the potential to be used for good (for example, to save lives through early treatment), or bad (for example, by unscrupulous firms to target vulnerable consumers), and the meaning or significance of any relationships it identifies will need to be interpreted. You should think carefully about your ethical issues and governance arrangements when using these techniques. The ethical considerations here are broad and cover (amongst other things): algorithmic bias[158] (the fact that algorithms may mimic the discriminatory biases that humans have), transparency and interpretability[159], and whether a social license for the use of big data exists.[160] In this vein, our forthcoming Manifesto for the future of applying behavioural science sets out a proposal for Data Science for Equity.

Additional tool: Incorporating new types of data

A further reason for excitement about machine learning, and why you may be interested in using it while Exploring, is that it is proving increasingly adept at analysing so-called 'unstructured data', like language.

Only a few years ago, quantitative data analysts were largely focused on the analysis of numbers, or at least words turned into numbers (such as survey responses coded from 1 to 10). Contemporary techniques, however, can use computers to find meaning in words, audio files, images, or other materials. As a result, they can find patterns associated with specific outcomes, such as whether a child described in a social worker's notes went on to be a victim of harm, or if a patient went on to develop cancer.

One area of promise is natural language processing (NLP). NLP is a field of study that looks at how we can help computers parse meaning from written or spoken language. One benefit of having the computer parse meaning from the written word is that computers can read far quicker than human beings and therefore can be used to understand patterns in very large bodies of text. In our early data science work, we used social workers' case notes to understand whether we can predict the outcomes of young people in social care. We used a technique called structural topic modelling[161] to identify groupings of words that commonly occur together and then looked to identify whether specific topics were likely to predict young people being readmitted to social services.

Critically, we then tried to understand whether these topics were meaningful to social workers through the use of semi-structured interviews. This helped put the findings from the data analysis into context and understand their concerns about using these algorithms to aid their decision making. Sadly, evaluations of the effectiveness of algorithms to predict social care outcomes have been mixed and given the high stakes discussions in social care, the decision of whether to use these techniques should not be taken lightly.[162] Natural language techniques are touted as potentially useful for social scientists and policymakers in the future, but they need to be evaluated and then implemented within a clear ethical governance framework.[163]

Conclusion

Anyone who is tasked with influencing a behaviour needs to understand the issues faced by the people who will be affected, and to involve those people in developing the policy, product or service. We have written Explore to share the techniques that we, at BIT, use to set our assumptions aside and become immersed in policy contexts. It is impossible to set aside all assumptions, but it is important to test them wherever possible. That is particularly true for policy makers, who often have to make decisions that directly affect the lives of many people.

The nature of a written document means that we have had to set out the four tools in a linear fashion. Our live projects, however, are often much more fluid and iterative. Things that you hear and see will suggest hypotheses that you want to test in wider data, and the reverse will often happen too. Intriguing findings from surveys and data analysis may send you back to interviews and field visits as you seek to build a deeper understanding and empathy for what seems to be going on. The everyday lived realities of policy and frontline services is generally far messier than our briefs and project plans would suggest.

Similarly, though the essence of the Explore phase is to be open to new understanding and insights, it should also throw up ideas for possible solutions. You should keep note of these ideas, but try to avoid seeking a solution too early, or letting your excitement for a particular intervention lead you to miss other insights or possibilities.
Now, armed with these tools and techniques, we hope you get out and Explore. Whether you are venturing into unfamiliar territory, or looking at familiar ground from a fresh perspective, it can be a hugely rewarding and productive experience. So, for your next project, we wholeheartedly encourage you to Explore the rich and varied terrain.

Authors

Alex Gyani, Cameron Tan, Karen Tindall & Rory Gallagher

Authors names are in alphabetical order

Acknowledgements

We would like to thank the many people who contributed to this document, which has been 4 years in the making. These are, in no specific order: David Halpern, Ale De Filippo, Matthew Barnard, Alex Sutherland, Mark Egan, Katy King, Nicky Quinn, Min-Taec Kim, Tiina Likki, Antonio Silva, James Lawrence, Nikola Spadina, Emily Cardon, Laura Moyce, Paul Calcraft, Belinda Jivapong, Michael Hallsworth, Matthew Holt, Jess Heal, Ravi Dutta-Powell, Michael Sanders, Evie Monnington-Taylor and Susannah Hume. We apologise for any omissions.

We would also like to acknowledge all of our colleagues at the Behavioural Insights Team, who have contributed to the thinking in this document and the case studies that we have included.

Finally, we would like to thank all of our partners and participants in all of our projects. All of the case studies listed in this paper would not have been possible without you.

Endnotes

1 Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). Qualitative research practice: A guide for social science students and researchers. Sage.

2 Hallsworth, M., & Kirkman, E. (2020). Behavioral insights. MIT Press.

3 We would refer readers who are interested in learning more about RCTs to Haynes, L., Goldacre, B., & Torgerson, D. (2012). Test, learn, adapt: developing public policy with randomised controlled trials. Cabinet Office-Behavioural Insights Team.

4 Of course, not all of our interventions are successful and we also aim to publish these

5 Atkins, L., Francis, J., Islam, R., O'Connor, D., Patey, A., Ivers, N., ... & Michie, S. (2017). A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implementation Science, 12(1), 1-18.

6 Rumelt, R. P. (2012). Good strategy/bad strategy: The difference and why it matters. Strategic Direction.

7 Thaler, R. H. (2015). Misbehaving. WW Norton.

8 Lewin, K. (1936) Principles of Topological Psychology. Translated by Heider, F. & Heider, G. New York: McGraw-Hill,

9 Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

10 This can be done in a formal 'Framework' method or otherwise. For an overview of the Framework method see: Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC medical research methodology, 13(1), 1-8. Framework methods allow you to use an explicitly theory driven way of coding your data in a behaviourally informed manner.

11 Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. Social and personality psychology compass, 10(9), 503-518.

12 All of this should of course be done with their permission.

13 Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. Perspectives on psychological science, 8(3), 223-241.

14 Heider. F (1958). The psychology of interpersonal relations. New York: Wiley

15 Wason, P. C. (1968). "Reasoning about a rule". Quarterly Journal of Experimental Psychology. 20(3): 273–281

16 Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. science, 211(4481), 453-458.

17 Wason, P. C. (1968). "Reasoning about a rule". Quarterly Journal of Experimental Psychology. 20(3): 273–281

18 Mannix, E., & Neale, M. A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. Psychological Science in the Public Interest, 6(2), 31–55.

19 Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. Behavioural Public Policy, 1(1), 54-86.

20 This is an aspirational goal, as it is not possible to do this entirely.

21 Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. The adapted mind: Evolutionary psychology and the generation of culture, 163, 163-228.

22 Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: a corrective strategy for social judgment. Journal of personality and social psychology, 47(6), 1231.

23 Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). Qualitative research practice: A guide for social science students and researchers. sage.

24 Gilbert, N. (2008). Research, theory and method. Researching social life, 2.

25 Cooper also coined the term 'cognitive friction' in his book. For more details read: Cooper, A (1999), The Inmates are Running the Asylum, SAMS.

26 Note that this does not mean a persona should be the average of the population on a set of variables, as this person often does not exist. However, personas should be sense-checked against the data to make sure that they represent a broad but realistic set of people in the situation you are exploring.

27 Meadows, D. H. (2008). Thinking in systems: A primer. Chelsea Green Publishing.

28 We note that there is a fair amount of controversy surrounding the idea of choice overload and note that meta-analytic data currently indicates that it is an established phenomenon but only under narrow circumstances. We found the following view very useful: Chernev, A., Böckenholt, U., & Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. Journal of Consumer Psychology, 25(2), 333-358.

29 If your research specifically focuses on a culture that you are not a part of, you should consider whether you are the best person to be doing this work. It might be the case that a researcher from that culture will be better placed. It might also be the case that you can do that work alongside someone from that culture.

30 Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care. data ran into trouble. Journal of medical ethics, 41(5), 404-409.

31 Kavallo, D., & Kahneman, D. (2003). Delusions of success: How optimism undermines executive decisions. Harvard Bus Rev. 81(7):56-63.

32 Data Futures Partnership. (2017). A Path to Social Licence: Guidelines for Trusted Data Use. Data Futures Partnership.

33 Klein, G. (2007). Performing a Project Premortem. Harvard Business Review 85(9), 18–19

34 When members of a group aim to make a decision, groupthink can occur. As they discuss possible courses of action, the collective desire to avoid conflict can sometimes lead to a consensus decision that overlooks important critiques or issues. See Janis, I. L. (1971). Groupthink. Psychology today, 5(6), 43-46.

35 Design Kit, IDEO.org (2016). The field guide to human centered design. IDEO

36 Ries, E. (2011). The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses. Currency.

37 In Eric Ries's Lean Startup this process is described as hypothesis validation. This will make anyone who has studied Philosophy of Science shudder. However, there is value in setting which answers will give you enough insight to move to the next stage of your Explore phase.

38 For a more detailed guide on how to do this, the following paper is very comprehensive and also has a behavioural focus: Atkins, L., Francis, J., Islam, R., O'Connor, D., Patey, A., Ivers, N., ... & Michie, S. (2017). A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implementation Science, 12(1), 1-18.

39 Design Council UK (2015). Design methods for developing services. *Keeping Connected Business Challenge.*

40 In this instance, you should make sure that you have a representative sample of the group that you are surveying. Ideally this is done by taking a random sample of your population of interest. This is rarely possible, so you might choose to use a different sampling method. These are covered in the Survey section.

41 NSW Behavioural Insights Unit (2020) Applying behavioural insights to drive public sector diversity

42 Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). Qualitative research practice: A guide for social science students and researchers. SAGE.

43 If you do want to understand disagreements, then make sure the focus group is a safe place for this. You should also try to avoid sensitive discussions in groups, as people may not feel comfortable sharing certain views or may feel embarrassed if they do.

44 Stewart, D. W., & Shamdasani, P. N. (2014). Focus groups: Theory and practice (Vol. 20). SAGE publications.

45 Edwards, A. L. (1957). The social desirability variable in personality assessment and research.New York, NY: Dryden.

46 Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. Journal of applied psychology, 84(5), 754.

47 Harper, H., & Hallsworth, M. (2016). Counting Calories. How under-reporting can explain the apparent fall in calorie intake. The Behavioural Insights Team.

48 Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American psychologist, 17(11), 776.

49 Percy, A., McAlister, S., Higgins, K., McCrystal, P., & Thornton, M. (2005). Response consistency in young adolescents' drug use self-reports: a recanting rate analysis. Addiction, 100(2), 189-196.

50 Hameed, W., Ishaque, M., Gul, X., Siddiqui, J. U. R., Hussain, S., Hussain, W., ... & Balal, A. (2017). Does courtesy bias affect how clients report on objective and subjective measures of family planning service quality? A comparison between facility-and home-based interviews. Open access journal of contraception, 9, 33.

51 Tversky, Amos; Kahneman, Daniel (1973). "Availability: A heuristic for judging frequency and probability". Cognitive Psychology. 5 (2): 207–232. doi: 10.1016/0010-0285(73)90033-9 ISSN 0010-0285

52 Shedler, J., & Manis, M. (1986). Can the availability heuristic explain vividness effects?. Journal of personality and social psychology, 51(1), 26.

53 Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. Field Methods, 18(1), 59-82.

54 Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: interviews and focus groups. British dental journal, 204(6), 291-295.

55 Light touch interviews may last from 25 to 45 minutes and in-depth interviews may range from 45- to 60 minutes.

56 Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). Qualitative research practice: A guide for social science students and researchers. SAGE.

57 We would still recommend having a note taker, even if you are recording the session. People can talk over each other in a focus group, so individual contributions can be difficult to hear on recordings. Some organisations will have facilities with multi-track recording setups to capture individual insights. However, this will mean that you will need to conduct the focus group in an environment which might be alien to your participants.

58 Janis, I. L. (1971). Groupthink. Psychology today, 5(6), 43-46.

59 Phillips, K. W., Liljenquist, K. A., & Neale, M. A. (2009). Is the pain worth the gain? The advantages and liabilities of agreeing with socially distinct newcomers. Personality and Social Psychology Bulletin, 35(3), 336-350.

60 Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). Noise: a flaw in human judgment. Little, Brown.

61 If the topic is sensitive, then you should also consider a private location.

62 The Five Whys. Retrieved from http://www.designkit.org/methods/66

63 Of course, if your participant looks upset, you should stop the interview and refer them to support.

64 Louw, S., Todd, R. W., & Jimarkon, P. (2011, April). Active listening in qualitative research interviews. In Proceedings of the International Conference: Research in Applied Linguistics, April.

65 If you record any calls or video calls, you should understand what your data storage obligations are and make sure that you meet them. It may be the case that you can't store recordings on personally owned devices or that the software that you are using to record video calls stores data on unsecured servers.

66 Surprising people by asking to record the session not only puts your participants into the uncomfortable position of declining if they do not want to be recorded, but will mean that the first question that they will ask themselves in the interview will be whether or not they trust you with the recording. This will put them on the backfoot and will make it less likely that they are open with you during the interview.

67 Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? Psychological Bulletin, 125(5), 591–600. https://doi.org/10.1037/0033-2909.125.5591

68 Street, J, Duszynski, K., Krawczyk, S. Braunack-Mayer, A. "The use of citizens' juries in health policy decision-making: a systematic review." Social science & medicine 109 (2014): 1-9.

69 Halpern, D. (2016). Behavioural insights and healthier lives. Melbourne: Victorian Health Promotion Foundation.

70 You can watch a video from the #NoFilterForum here.

71 You can learn more about the toolkit here.

72 Shen, F., & Han, J. (2014). Effectiveness of entertainment education in communicating health information: A systematic review. Asian Journal of Communication, 24(6), 605-616.

73 You can read more details on our findings here. The results from the impact evaluation of this intervention are forthcoming.

74 Fabbri, C., Rodrigues, K., Leurent, B., Allen, E., Qiu, M., Zuakulu, M., ... & Devries, K. M. (2021). The EmpaTeach intervention for reducing physical violence from teachers to students in Nyarugusu Refugee Camp: A cluster-randomised controlled trial. PLoS medicine, 18(10), e1003808.

75 Cooper, A. (1999). The Inmates Are Running the Asylum, SAMS.

76 Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. Psychological bulletin, 132(6), 895.

77 Ross, L., & Nisbett, R. E. (2011). The person and the situation: Perspectives of social psychology. Pinter & Martin Publishers.

78 Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. Psychological bulletin, 132(6), 895.

79 Rhoades, H., Wenzel, S. L., Rice, E., Winetrobe, H., & Henwood, B. (2017). No digital divide? Technology use among homeless adults. Journal of Social Distress and the Homeless, 26(1), 73-77.

80 Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC medical research methodology, 13(1), 1-8.

81 Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge: Cambridge University Press

82 Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. Annals of neurosciences, 20(4), 155. Translated by Ruger, H. A. & Busensis, C., E.

83 Loftus, E. F., & Palmer, J. C. (1996). Eyewitness testimony. In Introducing psychological research (pp. 305-309). Palgrave, London.

84 Shaw, J. (2020). Do false memories look real? Evidence that people struggle to identify rich false memories of committing crime and other emotional events. Frontiers in psychology, 11, 650.

85 Sedgwick, P., & Greenwood, N. (2015). Understanding the Hawthorne effect. Bmj, 351.

86 If you struggle with remembering some of your insights, there are tricks that you can use to improve your memory. (cf Dresler, M., Shirer, W. R., Konrad, B. N., Müller, N. C., Wagner, I. C., Fernández, G., ... & Greicius, M. D. (2017). Mnemonic training reshapes brain networks to support superior memory. Neuron, 93(5), 1227-1235) You can also use some techniques from Cognitive Interviewing when documenting your findings from any observations Fisher, R. P., & Geiselman, R. E. (2010). The cognitive interview method of conducting police interviews: Eliciting extensive information and promoting therapeutic jurisprudence. International journal of law and psychiatry, 33(5-6), 321-328.

87 This blog describes this work in more detail.

88 Using phones to capture data rather than clipboards can have benefits in that people are less likely to feel that they are being watched and therefore more likely to act naturally. However, it also means that people are less likely to know that they are being watched and you need to consider the ethical implications of this.

89 BIT (2018) Initial Findings: BIT review of the Benefit Change Notice for the Australian Energy Regulator.

90 Costa, E., & Halpern, D. (2019). The behavioural science of online harm and manipulation, and what to do about it. The Behavioural Insights Team.

91 Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. American Journal of Psychology, 113(3), 387-404.

92 Note that there are many companies offering technological solutions to this, we would strongly recommend sticking to doing this manually. A 2021 paper highlights the ethical and technical challenges in this domain. It notes that evaluations of facial recognition pilots in real world settings perform exceptionally poorly. Raji, I. D., & Fried, G. (2021). About Face: A Survey of Facial Recognition Evaluation. arXiv preprint arXiv:2102.00813.

93 Note that the term 'domestic violence' is being used here as it is the term used by the New South Wales government. However, the use of this term is contentious and the term 'domestic abuse' is now more commonly used in the English speaking world. (cf Hill, J. (2020). See what you made me do: Power, control and domestic abuse. Oxford University Press.)

94 Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. Journal of political Economy, 97(5), 1232-1254.

95 Merton, R. K. (1963). Bureaucratic Structure and Personality. In N. J. Smelser & W. T. Smelser (Eds.), Personality and social systems (pp. 255–264). John Wiley & Sons Inc. https://doi.org/10.1037/11302-024

96 NSW DPC Behavioural Insights Unit (2016) Behavioural Insights in NSW Update Report 2016. NSW Government Department of Premier & Cabinet.
97 NSW DPC Behavioural Insights Unit (2018) The Behavioural Insights Unit Report April 2018. NSW Government Department of Premier & Cabinet.
98 Kim, M.T., (2021). The impact of the 'What's Your Plan?' program on ADVO breaches and domestic violence (Crime and Justice Bulletin No. 242). Sydney: NSW Bureau of Crime Statistics and Research
99 Convery, E., Morse, A., Fung, B., Wodak, S., Powell, Z., Quinn, V., Taylor, M., Searle, B., & Vårhammar, A. (2020a). Work health and safetyof FDWs in the gig econimy. Report prepared for the Centre for Work Health and Safety (NSW).
100 Design Council UK (2015). Design methods for developing services. *Keeping Connected Business Challenge.*
101 Both CampaignMonitor and SurveyMonkey release benchmark response rates.
102 John, L. K., Acquisti, A., & Loewenstein, G. (2010). Strangers on a plane: Context-dependent willingness to divulge sensitive information. Journal of consumer research, 37(5), 858-873.
103 Dalton, D. R., Wimbush, J. C., & Daily, C. M. (1994). Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior. Personnel Psychology, 47(4), 817-829.
104 Roese, N.J. & Jamieson, D.W. (1993). "Twenty years of bogus pipeline research: A critical review and meta-analysis". Psychological Bulletin, 114, 363–375.
105 Warner, S L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." Journal of the American Statistical Association 60 (309). Taylor & Francis: 63–69.
106 Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological bulletin, 52(4), 281.
107 https://www.ukdataservice.ac.uk/
108 https://ipip.ori.org/
109 https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/ measuring-essential-skills/spectrum-database/
110 Many psychology textbooks will break this up into face validity and content validity. The former being based on the judgement of a layperson and the latter on the judgement of an expert.
111 In some cases you might also care about inter-rater reliability, which is the ability for a survey to have multiple people rate the same situation and have the same measurement. This is usually used when a survey requires some judgement on the part of the person administering it. For example, in BIT's recruitment process using Applied, we will double mark candidates' applications using a structured marking scheme. This allows us to see if people come up with the same score for each applicant and assess the extent of agreement between raters.
112 Meyer, V. M., Benjamens, S., El Moumni, M., Lange, J. F., & Pol, R. (2020). Global overview of response rates in patient and health care professional surveys in surgery: a systematic review. Annals of Surgery.
113 Keeter, S., Hatley, N., Kennedy, C., & Lau, A. (2017). What low response rates mean for telephone surveys. Pew Research Center, 15(1), 1-39.
114 Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. Computers in human behavior, 26(2), 132-139.
115 Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. Public opinion quarterly, 73(2), 349-360.
116 Liu, M., & Wronski, L. (2018). Examining completion rates in web surveys via over 25,000 real-world surveys. Social Science Computer Review, 36(1), 116-124.
117 Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. Public opinion quarterly, 73(2), 349-360.
118 Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of mobile versus PC web on survey response quality: A crossover experiment in a probability web panel. Public Opinion Quarterly, 81(S1), 280-306.
119 Liu, M. & Inchausti. N. (2017). Improving survey response rates: The effect of embedded questions in web survey email invitations. Survey Practice, 10(1), 2791.
120 Garner, R. (2005). Post-It® note persuasion: a sticky influence. Journal of Consumer Psychology, 15(3), 230-237.
121 Dillman, D.A. (2007) Mail and internet surveys, 2nd edition. John Wiley & Sons,Inc., Hoboken, NJ.
122 Sauermann, H., & Roach, M. (2013). Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. Research Policy, 42(1), 273-286.
123 Bosnjak, M., Neubarth, W., Couper, M. P., Bandilla, W., & Kaczmirek, L. (2008). Prenotification in Web-Based Access Panel Surveys The Influence of Mobile Text Messaging Versus E-Mail on Response Rates and Sample Composition. Social Science Computer Review, 26(2), 213-223.
124 Sauermann, H., & Roach, M. (2013). Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. Research Policy, 42(1), 273-286.
125 For an overview of research conducted by e-mail platforms (such as MailChimp) see https://coschedule.com/blog/best-time-to-send-email/
126 Chwalisz, C. (2021). The pandemic has pushed citizen panels online. Nature, 589(7841), 171-171.
127 World Health Organization. (2021). Voice, agency, empowerment: handbook on social participation for universal health coverage.
128 Sharp, A., & Anderson, K. (2010). Online citizen panels as an advance in research and consultation–A Review of pilot results. Commonwealth Journal of Local Governance, (6), 33-54.
129 Fishkin, J. S., & Luskin, R. C. (2005). Experimenting with a democratic ideal: Deliberative polling and public opinion. Acta Politica.
130 Institute for Government (2010) Pensions Reform The Pensions Commission (2002-6)

131 There are many online survey platforms and online randomisers, but we offer more than most through Predictiv because it also comes with support from a team of experts in behavioural insights, experimentation and policy analysis see www.predictiv.co.uk.

132 You can read more here.

133 An overview of some of our early Covid -19 work on Predictiv can be seen here. You can also analyse the data yourself!

134 You can read the report on this in English and Indonesian here.

135 You can read about our work with the Bank of England here.

136 Calcraft, P., (2022) Accelerating public policy research with easier & safer synthetic data

137 Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. O'Reilly Media, Inc.

138 Tukey, J. W. (1962). The future of data analysis. The annals of mathematical statistics, 33(1), 1-67. Quote taken from p 49.

139 Brody, H., Rip, M. R., Vinten-Johansen, P., Paneth, N., & Rachman, S. (2000). Map-making and myth-making in Broad Street: the London cholera epidemic, 1854. The Lancet, 356(9223), 64-68.

140 Wellcome Library, London. Wellcome Images images@wellcome.ac.uk http://wellcomeimages.org Map showing deaths from Cholera in Broad Street, Golden Square and the neighbourhood, from 19th August to 30th September 1854 1855 On the mode of communication of cholera / John Snow Published: 1855.

141 The following guide by the Data Viz Project will show you what you can do with the data that you have.

142 Harford, T. (2020). How to Make the World Add Up: Ten Rules for Thinking Differently About Numbers. Hachette UK.

143 Small, H. (2017). A Brief History of Florence Nightingale: and Her Real Legacy, a Revolution in Public Health. Robinson.

144 Tufte, E. R. (1983). The visual display of quantitative information Graphics Press. Cheshire, Connecticut.

145 Berinato, S. (2016). Good charts: The HBR guide to making smarter, more persuasive data visualizations. Harvard Business Review Press.

146 It is entirely possible that you may not need to use analytical methods to understand the data if your visualisations give you clear insights

147 Vigen, T. (2015). Spurious correlations: Correlation does not equal causation. Hachette books

148 For a fuller discussion of implementation issues see the interesting recent work of Al-Ubaydli et al (2019) and the implementation science literature more broadly.

149 Haynes, L., Goldacre, B., & Torgerson, D. (2012). Test, learn, adapt: developing public policy with randomised controlled trials. Cabinet Office-Behavioural Insights Team.

150 Jacobson, N. S., & Truax, P. (1992). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research.

151 This is the term for the framework that underpins most statistical testing used in the social sciences. There are other types of significance testing, but we will not cover them here. We would recommend reading: Cohen, J. (1994). The earth is round (p<. 05). American psychologist, 49(12), 997.

152 Yeturu, K. (2020). Machine learning algorithms, applications, and practices in data science. In Handbook of Statistics (Vol. 43, pp. 81-206). Elsevier.

153 For an idea of the complexity of models that can be produced, GPT-3 — a recent example of this approach applied to text data — has 175 billion parameters in its model. (Source)

154 This distinction is common, but by no means perfect. If you would like to know more, we recommend the following article.

155 Sometimes 'reinforcement learning' is offered as a third categorisation. See: Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and unsupervised learning for data science, 3-21.

156 For more examples of how we have used machine learning see Using Data Science in Policy, a report by the Behavioural Insights Team (December 2017)

157 This process is often repeated with different splits in what is known as "cross-validation".

158 Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. World Policy Journal, 33(4), 111-117.

159 Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76-99.

160 Shaw, J. A., Sethi, N., & Cassel, C. K. (2020). Social license for the use of big data in the Covid -19 era. npj Digital Medicine, 3(1), 1-3.

161 Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. Journal of the American Statistical Association, 111(151), 988–1003

162 Clayton, V., Sanders, M.T.,Schoenwald, E., Surkis, L. & Gibbons, D. (2020). Machine learning in children's services. What Works Centre for Social Care.

163 Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (Eds.). (2016). Big data and social science: A practical guide to methods and tools. crc Press.