The practical challenges of delivering field trials: how to reduce hassle and costs

Peter John

KCL

Basic proposition

- With observational data you just need to collect data
- With field experimental data you have to create an intervention in the real world from which to collect data:
  - **Two** interlocked tasks – need to work smoothly together
  - 'Double the trouble'
  - Little discussion in literature: 'you learn the hard way'.
- Practical work needs to be leveraged in a way that does not breach core assumptions – *not just project management*
- The world is complex and changing and you are weak:
  - limited resources, low importance, academics might not have the right skillset + we are too shy or bookish
- Do policy experiments offer a way forward?>>>>>>>>>>>>>>

Policy experiments

- A distinct form of RCT which involves the policy-maker delivering an intervention.
- Policy-maker randomises, researcher evaluates
- Different to participation experiments, such as GOTV (even these need some collaboration)
- Can be part of an official evaluation, but new model whereby done informally

More on policy experiments

- Efficient and low cost: no money need change hands
- Puts the policy-maker in command - if they get idea about randomisation and are enthusiastic about it > it works for them
- Can show that policy-makers are wrong > so they have to commit in advance to being tested, no covering up (but see Claire Dunlop on badgers)

Risks of policy experiments

- Policy experiments need enthusiasm and buy in if they are to become normal
- But RCTs hard to do, and might not always get off the ground, very complex > loss of enthusiasm?
- Research is in a weak position
- You can’t easily get them ethically approved
- They cock up
- Can be very costly for academics in terms of time, and and can be a challenge to produce academic outputs.

The Lanarkshire School milk experiment, 1930

- Tested whether providing free school milk could improve the health of children
- Experiment: 20,000 children in 67 schools participated, with half not getting free milk.
- Class assignments were done alphabetically, but teachers swapped the allocation causing the control students to do better –see ‘Student’ 1931 (W. S. Gosset)
- Shows challenges of working with policy-makers, especially at the delivery end
- But could have been avoided by cluster randomisation (by school) – simple design reforms often work

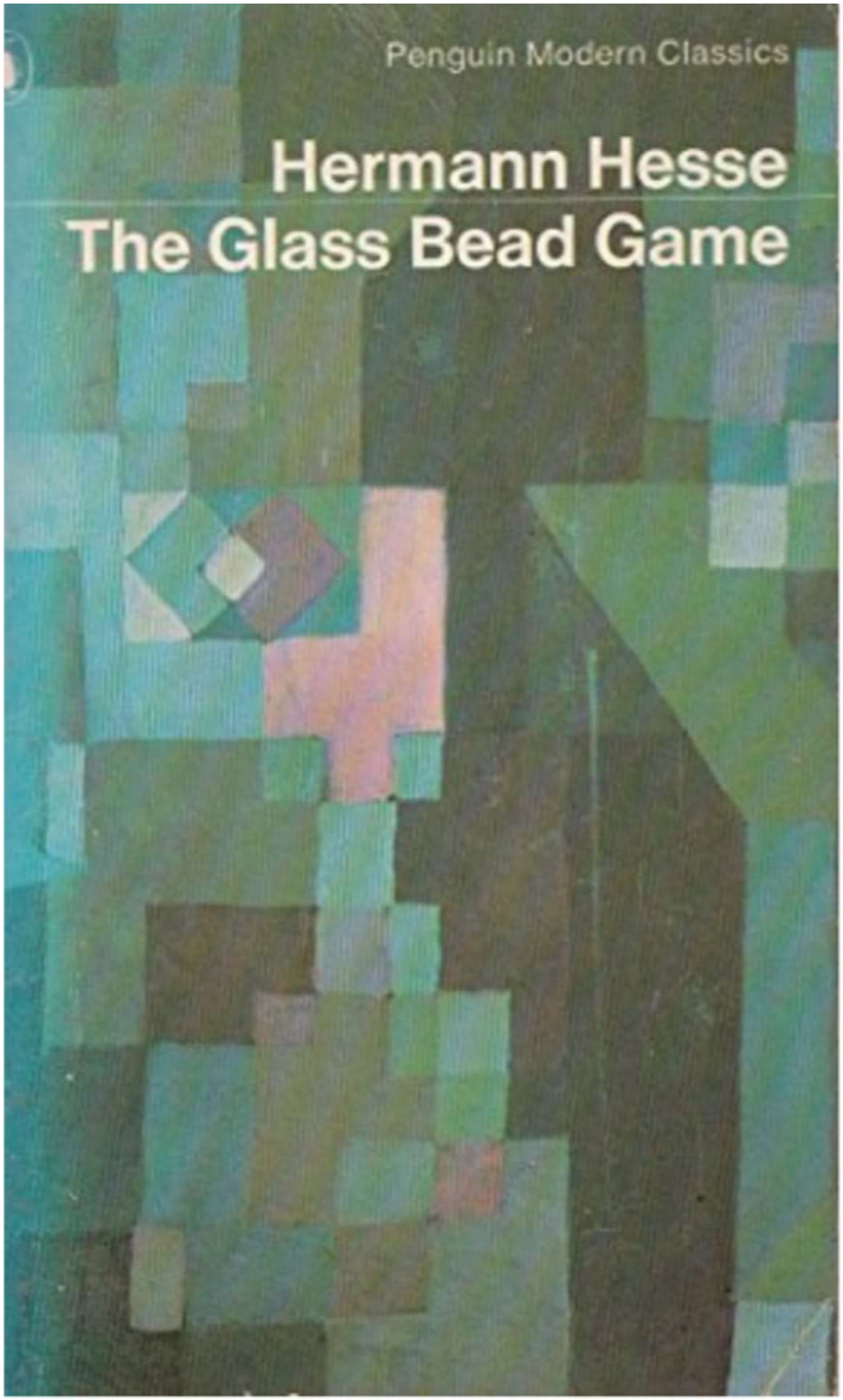
Peter John's cockup theory: ***if it can go wrong, it will go wrong***

- Policy-maker will lead you down the garden path then abandon you at the last moment – you get nothing
- Policy-maker screws up randomisation
- Outcomes can't be matched because of rolling data
- Government IT is hard to work with
- Another agency intervenes
- New people get involved with the experiment and veto
- Treatment is not delivered
- You always lose subjects --- you always lose subjects --- you always lose subjects --- you always lose subjects!


ROUTLEDGE STUDIES IN EXPERIMENTAL POLITICAL SCIENCE
Field Experiments in Political Science and Public Policy
Practical Lessons in Design and Delivery
Peter John

The ten steps of trial design

- *Step 1*: the assessment of fit for a trial
- *Step 2*: the recruitment of partners
- *Step 3*: the decision on the research site and sample
- *Step 4*: the determination of the treatments, is about testing the hypothesis.
- *Step 5*: the specification of the control condition
- *Step 6*: calculation of the sample size
- *Step 7*: agreement on the procedure for random allocation
- *Step 8*: collection of data on the participants or units
- *Step 9*: the assessment of ethical principles
- *Step 10*: the preparation of research plans


Penguin Modern Classics
Hermann Hesse
The Glass Bead Game

The Glass Bead game

- Had to drop this idea from the book as no one has heard of the book by Herman Hesse!
- Futuristic game but key idea is that it is on multiple levels which interlock so change one thing and everything changes – key to my scheme

Step 1: the assessment of fit for a trial

- A checking phase to ensure that the context is right. It encompasses a quick review of the other stages to see if there is a particular obstacle.
- Designed to ensure that nothing fundamental prevents a trial from taking place, such as the inability to deliver a treatment uniformly
- It is a fast-forward through the basic features of a trial, highlighting key to check out whether it is worth investing time and effort.
- This stage is particularly useful for the opportunistic trial, such as working with a policymaker
- Step 1 can help narrow options down or discard those with no hope or risky

Establish a question that can be answered by a trial

- Not all questions can
- Needs to be a good question – you do not know the answer, but also not too speculative
- Is the theory plausible (a mechanism you expect)? *Theory of change* (ToC)
- Can you intervene in a way that changes behaviour or outcomes (or attitudes) in a set time period?

Step 2: the recruitment of partners

- Also an early stage, especially for the more policy-orientated trial. A decision must be made to go with a partner
- This requires a choice, which again transports researchers to Step 1, because it is a chance to review all the options. Researchers need to pause at this stage as, once the choice of partner is made, it is hard to abandon later on. If the trial is not suitable, it needs to be discovered then.
- This step requires a set of practical skills needed to build a relationship with a partner, which often facilitates an agreement and may be the foundation for a good relationship as the trial is being implemented.
- How to recruit - long game
- Build trust
- Too quick, peak too early?
- Who to talk to – senior person, or more junior – both together?

Step 3: the decision on the research site and sample

- Linked to Step 2, but also involves choices that may be made with the partner about which sites and people to select into the experiment.
- It is another crucial step and it affects the external validity of the experiment as much as the selection of the partner.
- Very much linked to sample size questions

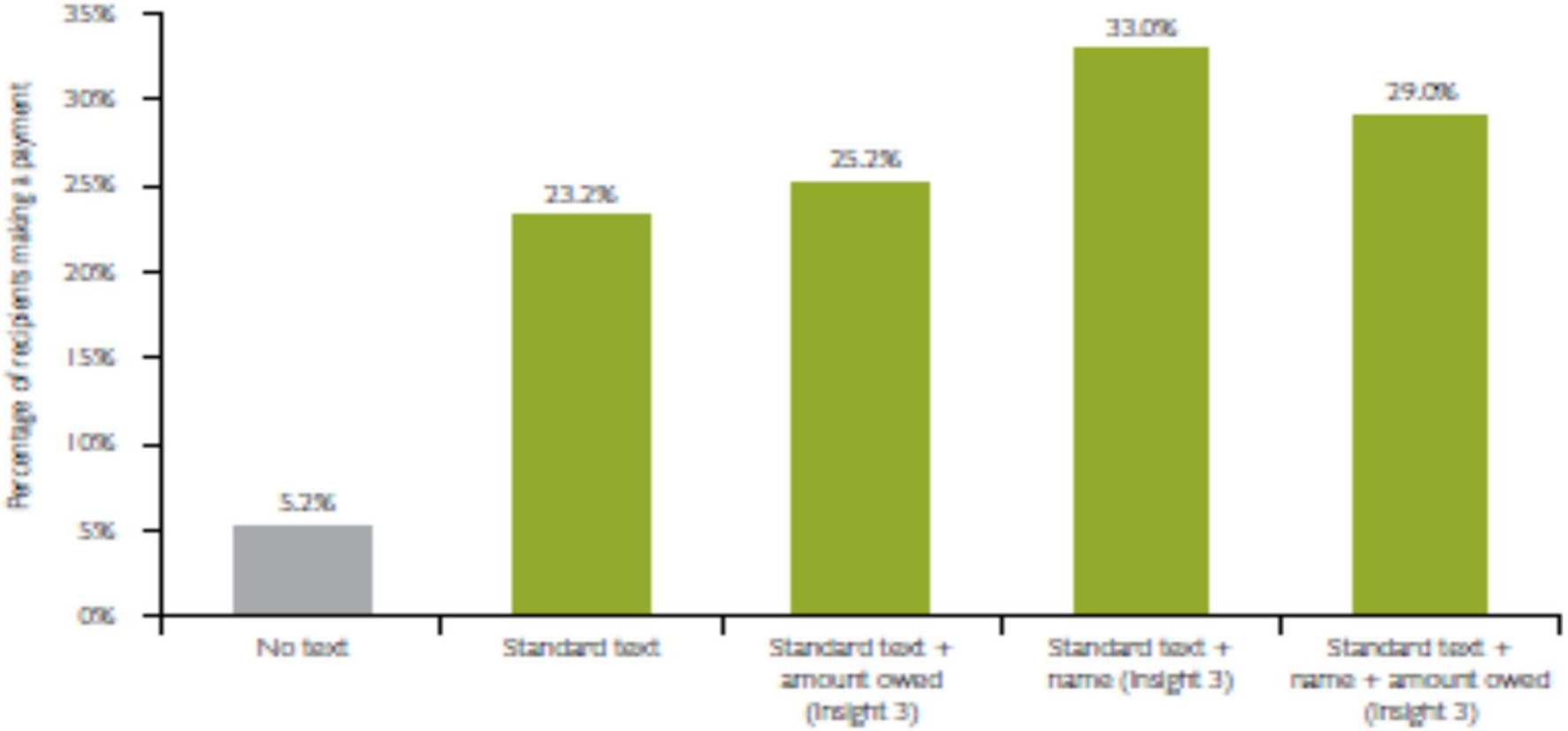
Step 4: the determination of the treatments

- Step 4 is about testing the hypothesis. It involves ensuring the treatment is an embodiment of the theory or hypothesis a researcher wants to test and only that, but it also involves making a series of practical choices that consider what is possible to do in the world and with the partner.
- A fault at this stage can be critical to the success of the experiment.
- Be ingenious!

A factorial design: texting experiment with court fines.

- “You have not paid your fine. Pay immediately or a warrant will be issued to the bailiffs. Call 03007909901 quote ref: XXXXXXX div XXX”
- “Joe Bloggs, you have not paid your fine. Pay immediately or a warrant will be issued to the bailiffs. Call 03007909901 quote ref: XXXXXXX div XXX”
- “You have not paid your fine of £XXXX. Pay immediately or a warrant will be issued to the bailiffs. Call 03007909901 quote ref: XXXXXXX div XXX”
- “Joe Bloggs, you have not paid your fine of £XXXX. Pay immediately or a warrant will be issued to the bailiffs. Call 03007909901 quote ref: XXXXXXX div XXX”

Results


Percentage of recipients making a payment
35%
30%
25%
20%
15%
10%
5%
0%
5.2%
23.2%
25.2%
33.0%
29.0%
No text
Standard text
Standard text + amount owed (Insight 3)
Standard text + name (Insight 3)
Standard text + name + amount owed (Insight 3)

Step 5: the specification of the control condition

- Is tied up with the previous step: it is the counterfactual from which the researcher makes an inference.
- It is not usually included in lists of this type, but is stressed because it tends to be forgotten about, although what happens to this group is important for making an inference about the effect of the intervention.

Step 6: calculation of the sample size

- Is one of the questions frequently asked by practitioners
- However, it is tricky to figure out in practice because there are many unknowns, such as expected effect size and the standard deviation. Researchers have to do their best using sample size calculators
- Scenarios a good idea

Step 7: agreement on the procedure for random allocation

- a task researchers need to plan for, as it is so crucial for the success of the trial and intersects with what is possible in the practical environment in terms of sharing data. The kind of randomization needs to be agreed upon as well.

Step 8: collection of data on the participants or units

- Data collection entails practical decision based on the availability of potential data and more general social science questions about what needs to be measured and whether what is collected represents the concepts being evaluated.
- Researchers have to anticipate any threats to the data, such as attrition, at this step.

Step 9: the assessment of ethical principles

- Ethics needs to be woven into the consideration of the other steps, but also deserves a phase of its own when the trial is beyond its first scoping phase and it is clear what treatments are to be provided.

Ethical issues

- All research needs to be guided by ethical issues – consent, avoidance of harm, respect for participants (e.g. right to withdrawal, privacy), maintain safety
- Trials often thought to be vulnerable because of randomisation may ‘deny’ benefits or impose costs (in practice these happen anyway)
- Also may involve deception that needs to be justified

Example

- Glewwe, P., Park, A., and Zhao, M. 2012. Visualizing development: Eyeglasses and academic performance in primary schools in China. Center for International Food and Agricultural Policy Research, University of Minnesota, Working Paper WP12-2 (Jan.)

China experiment

- Would wearing corrective eyeglasses might enable sight-defective kids to perform better at school?
- Thousands of schoolchildren—up to one half of the 19,000 student sample—were randomly chosen to be experimental “controls” in the eyesight experiment
- Students randomly selected to serve as “controls” were not supplied with corrective eyeglasses nor any other eye care, however sight defective and prepared they were to benefit from the best practice treatment, if given.
- Students in the control group were recruited and followed and tracked, same as the “treatment” group, but under no circumstance were the “controls” to be given corrective eyeglasses

Step 10: the preparation of research plans

- Different kinds of plans: external funding, project plans, ethical plans, delivery plans
- Registration more common: designed to prevent fishing
- See EGAP website

The nine threats

- Problem 1: Loss of participants
- Problem 2: Failed or compromised randomization
- Problem 3: Treatment not given or modified
- Problem 4: the control group gets treated
- Problem 5: interference between treatment and control
- Problem 6: unavoidable confounders
- Problem 7: poor quality data
- Problem 8: Loss of cooperation of partners
- Problem 9: unexpected logistical challenges and expenses

1: loss of participants

- Critical for so many experiments and is the reason so many are underpowered. It happens to a certain degree in every experiment - even ones where the units are fixed, like streets or villages as some units end up getting excluded.
- The question is whether there are enough people or units left to be able to make a statistical inference.

Problem 2: failed or compromised randomization

- Can happen because of a lack of cooperation between partners.
- Partners do not understand randomisation
- Some randomisations hard to implement – e.g. daily, or face to face with clients
- What this means is that experimenters need to be fastidious or obsessive at this stage.

Problem 3: treatment not given or modified

- This is very common and requires an experimenter and policymaker to monitor an experiment constantly.
- Basically a delivery problem by agencies

Problem 4: the control group gets treated

- Interference is another real world problem faced by trials, but it can be avoided through careful control over the experiment.
- Caused by poor implementation and human agency in an experiment
- Also enthusiasm from the recruited can assist this

Problem 5: interference between treatment and control

- Does happen but many researchers design their trials to try to avoid this, and there are ways to analyze the results to cope with the problem.
- Really about contact, information flows as so many experiments informational
- See Day 3 for how to deal with

Problem 6: unavoidable shocks

- Where external events affect outcomes both in treatment and control, wiping out any treatment effects
- Cannot be controlled by the researchers but usually is not so severe as to blow an experiment off course.
- Another government policy counts
- Natural events - e.g. snowstorm – you can stop and then restart the experiment

Problem 7: poor quality data

- Is very common, especially in policy experiments, and reflects the nature of official systems for collecting data and the difficulty of specifying in advance exactly the data needed and procedures for its transfer.
- Need to find out what data means - e.g. what is a debt

Problem 8: loss of cooperation by partners

- Is a factor that can undermine an experiment. In general, this tends not to happen once an experiment has been agreed upon, but there can be a cooling of the relationship as the partners experience the delivery of an experiment and realize its cost.
- For HMRC - about internal relations but can work in the same way

Problem 9: unexpected logistical challenges and expenses,

- Is a residual category, and affects the other problems, but happens a lot of because of the challenges, complex interaction between the intervention and the research, and costs of delivering an experiment.

Conclusions

- Really about adaptive design, enough planning and flexibility to deal with interactions between design features – back to the glass bead game
- Building in costs and budgeting for them
- Expendable trials – you have other projects also able to deliver
- Avoiding early euphoria
- Playing the long game – embracing sunk costs!

RCTs targeted to under-registered groups

Our project

- Uses randomised controlled trials to find out how to increase the electoral participation of low-voter-registration groups funded by the Joseph Rowntree Reform Trust as part of its UK Democracy Fund /).

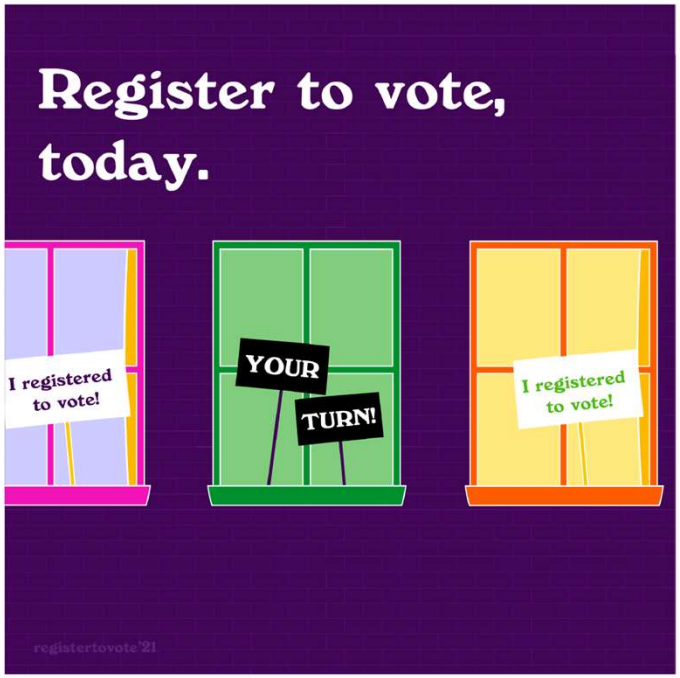
Messaging: Cognitive vs. non-cognitive mobilisation

- Almost all turnout and registration campaigns focus on increasing citizens' motivation to vote. An
- Empowerment is a cognitive strategy

- BUT:
- 'there are a great many citizens who fail to register/vote even though they intend to do so. What they need is resources that help them get through the process.' (Holbein and Hillygus, 2020, p.2).
- Need non-cognitive strategies, giving people resources to complete a a task they want to do

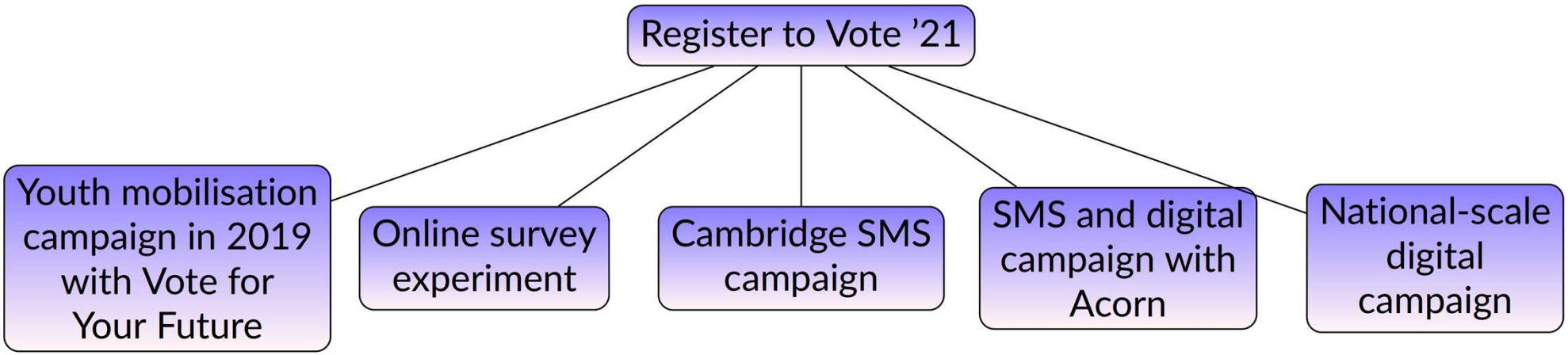
Cognitive Mobilisation ads

Dynamic norm ads



Anti-sludge ads

Our project design



Online survey experiment

- Outcome: intention to register to vote and vote
- Testing all the treatment arms against control group
- Testing cognitive vs. noncognitive

	Unregistered GE	Registered GE	Unregistered LE	Registered LE
Intercept	7.333^{***}	8.621^{***}	5.750^{***}	7.717^{***}
	(0.385)	(0.122)	(0.453)	(0.161)
Treatment	−0.379	−0.052	0.551	0.346^{*}
	(0.414)	(0.134)	(0.483)	(0.173)
R^2	0.002	0.000	0.003	0.002
Adj. R^2	−0.000	−0.000	0.001	0.002
N	478	1983	478	1983

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table: Overall effect

Trial 2: A SMS trial with Cambridge

Each participant is randomly assigned to one of three conditions in equal proportion: 1) a non-cognitive message, 2) a cognitive message, 3) a pure control of no message

	Register
Intercept	0.012
	(0.009)
Cognitive	0.098***
	(0.026)
Noncognitive	0.092***
	(0.025)
R^2	0.029
Adj. R^2	0.025
N	493

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table: Cognitive vs. noncognitive