

Can experiments inform regulators?

Steffen Huck
UCL Economics

State of play

Some very general lessons drawn from fairly abstract experiments (Krause, Kroger, Potters 2004).

Use of experiments to inform specific policies is very rare.

Market design more common (specifically auctions, including UK 3G spectrum auction).

Australian government agencies have been employing experimental studies to inform water policy.

The rise of behavioral economics

has triggered broad interest among government agencies.

That's no surprise because much of BE is so intuitive
(sometimes deceptively so).

UK and EU agencies commissioning behavioral analysis and
experimental studies.

What experiments can offer beyond generic advice

Detect problems consumers have in decision making

Identify the root causes of these problems

Road test remedies designed to help consumers

Test firms' responses to regulations, with simulated or actual buyers

Types of experiments

Lab has advantage of high levels of control and low costs

Field has advantage of realism and broader subject pools

The key issue of external validity

Lab can have better external validity than field

There are no general rules on which experiments have good or bad external validity

External validity is often highly asymmetric and depends on actual results:

For example, if "smart" student subjects struggle with "simple" decision problems, the general population will struggle with more complex versions of these tasks as well.

On the other hand, if "smart" student subjects are good at solving "simple" tasks, we learn very little about what this may mean for real life.

Tradeoffs in experimental design

Simplified tasks give great external validity if student sample struggles but one can be left empty-handed if they don't

Realism can increase external validity but also increase noise such that larger samples would be needed to obtain significant results

In general, good experimental design relies on a lot of judgment and nothing is for sure

Examples of recent work

detect problems consumers have with framing/presentations of prices and identify root causes (OFT)

roadtest measures to increase price transparency in telecom market (ofcom)

compare different switching processes in telecom markets (ofcom)

The impact of price frames on consumer decision making

Steffen Huck, Brian Wallace

What are price frames?

A price frame refers to the way a price is presented (“framed”).

We look at

- * straight per-unit prices *“This good costs £1 per unit.”*
- * drip pricing *“... plus shipping ... plus handling.”*
- * The rhetoric of sales *“Was £2 is now £1.”*
- * complex pricing *“3 for 2”*
- * baiting *“£1 while stocks last.”*
- * time-limited offers *“£1 only today.”*

Do price frames matter?

Economic theory says no.

Yet, we see sellers spending money on altering price frames.

Why would they if consumers would comply to theory?

Existing evidence

Not much. If sellers know about these matters, they also know how to hide it.

Best existing evidence probably on “partitioned pricing” shown to be effective on ebay (Hossain and Morgan 2006).

Scattered evidence on some of the other practices but there is no single study trying to analyse these frames in one and the same environment.

What we do

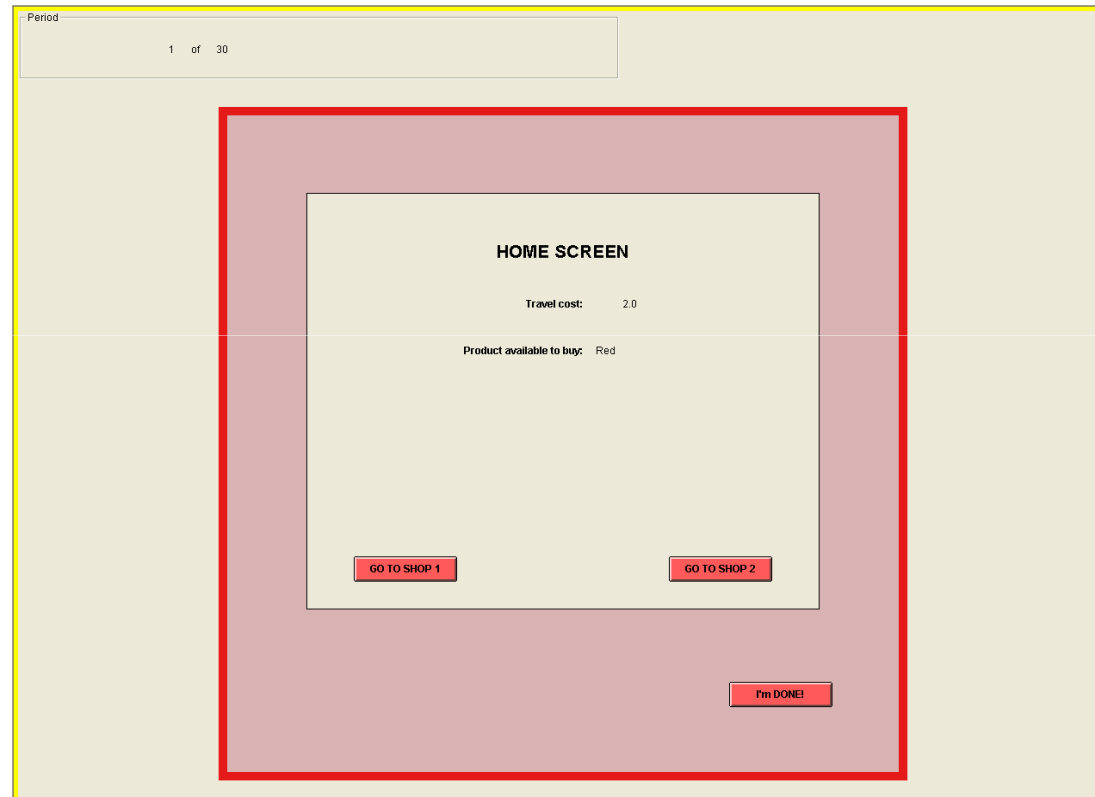
Design a laboratory experiment where (student) subjects are exposed to the different frames.

Real money is at stake.

Subjects are endowed with a “payoff function” that maps units of a good purchased into earnings. For example, 120 for first unit, 80 for second, 20 for third, 10 for the fourth.

There are two shops and search is costly.

The home screen



A shop with straight prices

Period

SHOP 1

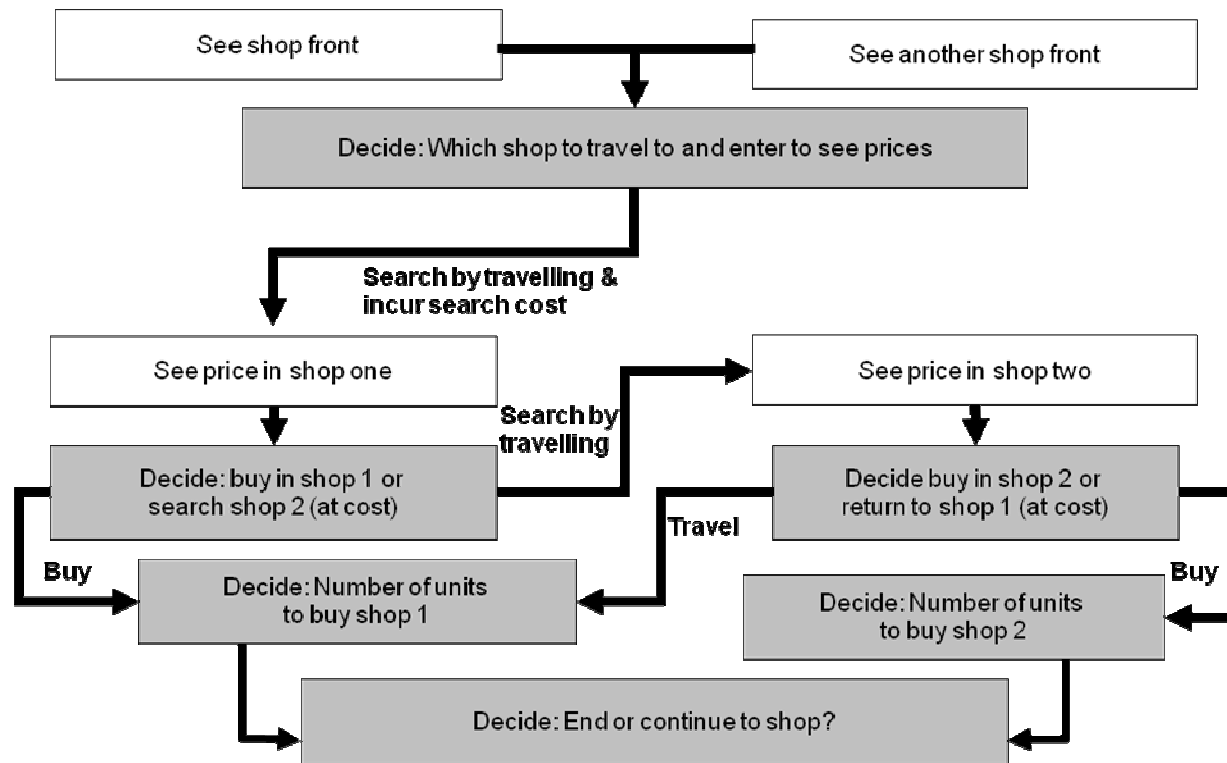
Price of product (per unit): 105

Units to buy:

[Buy](#)

[Go home!](#)

The consumer's problem



Experimental design

Shops draw prices at random from $[60, 120]$.

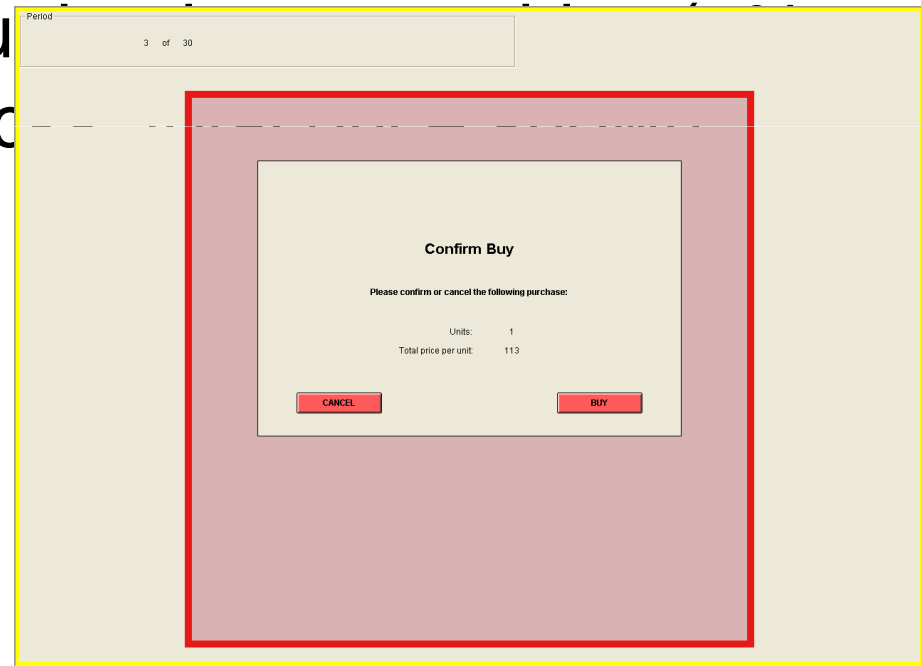
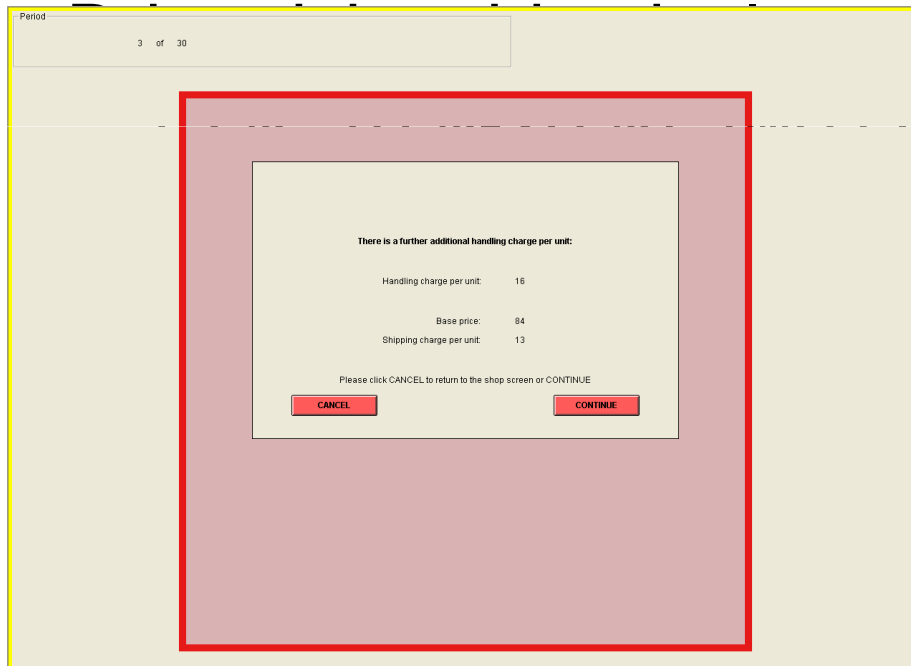
Three levels of search costs.

Four different ways of scaling payoffs (“four different types of good”)

166 subjects each of which confronts three different frames where all 10 combinations of 3 out of 5 are implemented. Frames in random order.

Pre-experimental understanding test, post-experimental IQ and personality tests, questionnaire

Implementation of frames



Implementation cont'd

Sales: identical unit prices, “former price” between current and max price. Implied % discount shown as well.

Complex: identical unit prices but “3 for 2” offer, ie, third unit is free.

Time-limited offers: Price just now, on return visit new price is drawn from same distribution, so price can go up but also down.

Baiting: Shops advertise price under generic “while stocks last” warning. True prices determined as before but if true price > 72 , the shop advertises a lower price half the time.

Optimal search

Identical for baseline, sales, and drip pricing:

Buy at shop 1 if price is below a certain cutoff, continue search otherwise. At second shop, return only if price difference compensates for search costs.

Different cutoffs for complex and time-limited.

Non-trivial selection of shops under baiting.

Otherwise identical to baseline.

This was risky

Take drip pricing. It's just two mouse clicks that separate drip pricing from the baseline. Can this have an effect?

In a highly selected student population?

What if it doesn't? Then we learn nothing.

But, hey!, what if it does?

Baseline results

Subjects can do this rather well but there is a tendency to search too much.

		Actual choice					
Behaviour at the first shop		0	1	2	3	4	Total
Optimal choice	0	804	93	25	1	3	926
	1	83	99	11	1	1	195
	2	119	14	410	3	4	550
	Total	1006	206	446	5	8	1671

Baseline cont'd

		Actual choice					
At the second shop ⁰		1	2	3	4	Total	
Optimal choice	0	144	17	2	0	0	163
	1	24	313	31	3	3	374
	2	2	18	234	3	3	260
	Total	170	348	267	6	6	797

Drip pricing has an effect

		Actual choice					
At the first shop		0	1	2	3	4	Total
Optimal choice	0	274	68	28	0	5	375
	1	25	39	7	0	0	71
	2	39	7	137	0	6	189
	Total	338	114	172	0	11	635

Oversearch is eradicated. Instead there is now undersearch.

Drip pricing

With the decomposition of the price into a base and two drips (just two mouse clicks!) consumers are more likely to buy.

Not only do they buy at prices where they exhibit a tendency to search too much under straight prices, *they even buy where they should continue to search.*

This wipes out some 25% of consumer welfare.

All frames have adverse effects

Even the innocuous sales frame!

Mostly, they trigger more search errors. Consumers are too willing to buy something at the first shop they visit.

An interesting exception is time-limited offers where consumers do not return often enough to the first shop when they face high prices at the second.

What is driving this?

Let's consider drip pricing.

Consumers display a higher willingness to pay at the first shop.

Cannot stem from sunk cost arguments (that might play a role in real life).

Only reasonable explanation is *loss aversion*. The moment consumers see low price they imagine buying the good at this price. Aborting the purchasing process would then generate a loss relative to the new reference point. This increases willingness to pay.

The role of search costs

With higher search costs there tend to

- be fewer search errors (consumers should buy more often at the first shop!)
- but more purchasing errors (consumers apparently want to ex post justify the incurred costs – an incarnation of the sunk cost fallacy).

Does learning help?

Yes, but to a limited degree. Learning improves behaviour but learning stops too early.

There is no learning under time-limited offers.

And what about IQ?

Measured on scale between 0 and 12.

1 IQ point is roughly worth 4 periods of learning.

Results: The big picture

Frame	More p err	More s err	C welf loss	Sold units	Sales	1 st shop +	Behavioural biases	Learning
Baseline				1.58	125			
Drip	Yes, a lot	Yes	Large	1.59	124	Yes	Endowment effect/loss aversion/maybe sunk cost fallacy	Yes
Time-ltd	Yes	Yes	Medium	1.59	121	Yes	Cognitive errors/maybe sunk cost fallacy	No
Baiting	Yes	Yes	Medium	1.61	124	Yes, strongly	Endowment effect/loss aversion/sunk cost fallacy	Yes
Complex	No	No	Small	2.73	221	Yes	Cognitive errors/sunk cost fallacy	Yes
Sales	Yes	Yes	No	1.59	125	No	Cognitive errors/sunk cost fallacy	Yes

External validity

Effects will be worse in general population.

Effects of real-life frames will also be worse as they have probably been optimized; also prices will have been adjusted.

Caveat: we measure the pure effects of frames, not of true offers

What about different markets? For some products it's easier to imagine ownership than for others. What with very high search costs?

What about firms?

The reported effects make it attractive to be the first shop a consumer visits.

Overall effect on industry profits could be negative!

Sellers could establish a reputation for not using such frames.

However, given the role of cognitive ability, it could be those consumers that suffer most under price framing that are least likely to perceive the advantages of a reputation for straight pricing ...

Price transparency in telephony

Charlotte Duke, Steffen Huck, Brian Wallace

Background

Consumers have little knowledge about true costs of non-geographical numbers (worse for calls from mobiles).

Ofcom is considering intervening in the market.

We roadtest different interventions.

<http://www.ofcom.org.uk/research/telecoms/reports/experiments/>

Basics

This is an exploratory study.

There is no guiding model (“game”).

We need to think about what is the essence of the problem.

A theorist would do the same (in the arm chair) but we get to observe actual behaviour.

Making decisions about calls

We focus on two type of decisions:

1. Does it pay to make calls to one particular number? (Ex. Call your bank/utility company or check the internet.)
2. Which number shall I call? (Ex. Choose between different directory enquiries/ticket agents.)

The baseline

We try to create a world that captures the essence of the problem.

There are different “tasks” that can be completed by making phone calls.

Making a phone call means simply pressing a button on a computer screen.

Through completing tasks subjects earn money but they have to pay the calling charges.

Call charges

For each task call durations are drawn from the same distribution. So numbers differ only in the per-minute charge.

The per-minute charges can be found out through search.

Search is conducted through clicking on a button and paying money for this (which adequately mirrors the time costs of such search outside the laboratory).

Phone bills

Subjects complete 14 cycles of 9 tasks, first in the baseline, then in an intervention.

After each cycle they receive an itemized phone bill from.

A decision screen

Cycle 1 of 14

EXIT

Task Information

Task number: 6
 Task premium: 120
 Type of task: SELECTION

Number	Cost per minute (p)	Click to call
16	Search	Call
26	120	Call
36	Search	Call
46	40	Call
56	Search	Call
66	Search	Call
76	30	Call
86	Search	Call

Design

Baseline replicating current situation where search for price information is rather costly.

Four interventions

PCA-exact, PCA-max

Short codes, Price lists

Two environments

Landline, Mobile

Exact Pre-call announcements

When subjects press the call button they get to see the charge per minute and can abort the call if they want to. There are slight time delays to reflect the “annoying” aspect of having to wait for these announcements.

Subjects can opt out of the PCA scheme (and opt back in) at any point in time.

Pre-call announcements with max charges

When subjects press the call button they get to see the maximum charge per minute for that service and can abort the call if they want to. There are slight time delays to reflect the “annoying” aspect of having to wait for these announcements.

Subjects can opt out of the PCA scheme (and opt back in) at any point in time.

Short codes

We capture short codes by drastically reducing search costs (94% for landline, 98% for mobiles).

Price lists

Ordered price list on phone bill.

But order is by number not by task.

Price lists on the bill

Cycle 1 of 14 EXIT

Number	Charge (pence per minute)	Number	Charge (pence per minute)	Number	Charge (pence per minute)
11	30	14	8	17	100
21	23	24	8	27	120
31	29	34	5	37	90
41	27	44	8	47	100
51	30	54	7	57	70
61	30	64	8	67	80
71	29	74	7	77	90
81	33	84	7	87	120
91	31	94	9	97	100
12	8	15	16	18	33
22	6	25	12	28	24
32	4	35	16	38	32
42	5	45	17	48	34
52	7	55	16	58	32
62	8	65	21	68	42
72	8	75	19	78	37
82	8	85	17	88	33
92	8	95	13	98	27
13	45	16	10	19	40
23	50	26	5	29	40
33	55	36	30	39	45
43	60	46	40	49	30
53	40	56	50	59	25
63	45	66	20	69	50
73	50	76	15	79	55
83	55	86	60	89	35
93	65	96	25	99	20

HIDE CALL CHARGES

Landlines and mobiles

Compared to the landline the mobile environment is characterized through higher charges, greater variance in charges, and bigger search costs.

Procedures

Pilots with 60 subjects

4x2 between-subject design with 211 subjects from
ELSE subject pool

Experiments lasted approx 150 minutes

On average, subjects earned £23.65

How we look at the data

We can compute what an omniscient subject who would know all the call charges and average call durations would do and what this omniscient subject would earn.

We can then compare our subjects' earning against this standard.

That is, we can compute the welfare losses in each environment as a percentage of an ideal scenario.

Welfare with and without search costs

In the field search costs would be difficult to measure.

Field investigation would probably just focus on monetary components, the benefits from making calls vs their costs.

So we examine welfare with and without search costs.

Landline results

Table 1: Performance in the landline environment

Landline	Number of participants	Total Welfare	Welfare 1st half	Welfare 2nd half
Baseline	113	-62.5%	-99.1%	-25.8%
PCA-exact	30	25.4%	6.7%	44.0%
Short codes	30	12.4%	4.1%	20.7%
Price list	25	-16.1%	-47.8%	15.5%
PCA-max	28	-13.9%	-43.1%	15.4%

Landline results w/o search costs

Performance in landline environment (excluding search costs)				
Landline	Number of participants	Total Welfare	Welfare 1st half	Welfare 2nd half
Baseline	113	-23.9%	-32.1%	-23.9%
PCA-exact	30	32.8%	21.4%	44.1%
Short codes	30	17.2%	13.0%	21.4%
Price list	25	-4.7%	-25.0%	15.5%
PCA-max	28	3.4%	-9.6%	16.5%

Mobile results

Table 1: Performance in the mobile phone environment

Mobile N	Number of participants	Total Welfare	Welfare 1st half	Welfare 2nd half
Baseline	98	-55.7%	-94.5%	-16.9%
PCA-exact	26	27.4%	16.2%	38.6%
Short codes	26	13.8%	0.9%	29.5%
Price list	26	11.8%	-11.7%	35.3%
PCA-max	20	-9.4%	-33.0%	14.1%

Mobile results w/o search costs

Table 4: Performance in mobile environment (excluding search costs)

Landline	Number of participants	Total Welfare	Welfare 1st half	Welfare 2nd half
Baseline	98	-10.7%	-15.5%	-6.0%
PCA-exact	26	29.7%	20.7%	38.7%
Short codes	26	17.7%	5.0%	30.4%
Price list	26	23.5%	11.4%	35.5%
PCA-max	20	10.3%	4.0%	16.7%

Ranking the interventions

PCA-exact > Short codes >> Price lists/PCA max

Same ranking in both landline and mobile environment suggests robustness.

Ranking holds through the entire IQ distribution.
(High IQ subjects just learn faster to make the most of the interventions.)

There are many other things we can do with the data

There are more slides with observations on searches, bill shock, opt outs and IQ.

But perhaps it's better for now to ask ourselves the key question.

What about the external validity of these results? Can we really learn something from this about the real world?

The key insight

If the design *biases in favour of* an intervention, then the experiment is only informative if the intervention doesn't do well.

If the design *biases against* an intervention, then the experiment is only informative if the intervention does do well.

External validity: How robust are results?

Design biases *against* interventions which will be even more useful in a dynamic changing environment.

Design biases *in favour* of price lists (because they are so short) and PCA-max (because max is actual price in some tasks).

=> If anything, we underestimate the true advantage of PCA-exact (and short codes).

External validity: The role of the student sample

A more representative sample would do worse in all environments.

There is no reason to expect this would affect the relative rankings.

Relative quantitative effects of interventions that require little cognitive effort may be larger in representative sample. Again, this would favour PCA-exact.

The key results

Interventions do work

Good performance of PCA-exact with opt-out and short codes

Bad performance of PCA-max (unsurprising) and price lists (much more surprising and illustrating the role bounded rationality)

Everybody in IQ distribution benefits from interventions.

Other findings

- IQ matters for performance
- High IQ gain immediately from interventions, low IQ need some more time but benefit as well
- Despite substantial parameter differences between landline and mobile overall picture quite similar
- Opt-outs are important

Wider lessons

Lab experiments are very useful in identifying the *relative performance* of different policies.

Lab experiments are particularly good in identifying interventions / policy measures that will fail.

When do experiments?

Early on! To get rid of the bad ideas.

Then check feasibility of a field experiment that focuses on policies that survived the lab!

Supplement field experiment with other perhaps more qualitative studies if in doubt.

General recommendations for design procedure

Design choices much harder *and* more important in the absence of a model.

Carefully check availability of theoretical models and empirical findings that can inform design.

If not, do take your time ...

Some general conclusions

External validity becomes obviously much more important than in experimental econ literature.

Biased design and non-representative samples can be good because of asymmetries in external validity!

Experiments are just like theory.