# Full disclosure: a round-up of FCA experimental research into giving information

Occasional

Paper 23

November 2016

Encouraging debate among academics, practitioners and policymakers in all aspects of financial regulation.

> Financial Conduct Authority www.fca.org.uk

## FCA occasional papers in financial regulation

#### The FCA occasional papers

The FCA is committed to encouraging debate on all aspects of financial regulation and to creating rigorous evidence to support its decision-making. To facilitate this, we publish a series of Occasional Papers, extending across economics and other disciplines.

Occasional Papers contribute to the debate on specific issues relevant to the FCA's work. The main factor in accepting papers is that they should make substantial contributions to knowledge and understanding of financial regulation. If you want to contribute to this series or comment on these papers, please contact Peter Andrews or Kevin James at peter.andrews@fca.org.uk and kevin.james@fca.org.uk

#### Disclaimer

Occasional Papers contribute to the work of the FCA by providing rigorous research results and stimulating debate. While they may not necessarily represent the position of the FCA, they are one source of evidence that the FCA may use while discharging its functions and to inform its views. The FCA endeavours to ensure that research outputs are correct, through checks including independent referee reports, but the nature of such research and choice of research methods is a matter for the authors using their expert judgement. To the extent that Occasional Papers contain any errors or omissions, they should be attributed to the individual authors, rather than to the FCA.

#### Author

Laura Smart

#### Researchers

Paul Adams, Matteo Aquilina, Robert Baker, Will Brambley, Alessandro Nava, Sumedha Pathak, James Ridgewell, Helena Robertson, James Shafe, Laura Smart, Dom Suckling, Roisin Wilson and Qamar Zaman.

#### Acknowledgements

We would like to thank Daniele Nosenzo (University of Nottingham) and Heather Kappes (London School of Economics and Political Science) for their academic review, and Darragh Kelly, Ben Guttman-Kenney, Jeroen Nieboer, Alice Ciccone, James Korolus and Matt Clark, for their help with statistical analysis and review. Finally, we would like to take this opportunity to thank all of the organisations and FCA staff who worked with us to carry out the trials and experiments in this paper and in our previous papers. We sincerely value the collaboration and commitment of those we worked with who were instrumental in making this happen.

All our publications are available to download from www.fca.org.uk. If you would like to receive this paper in an alternative format, please call 020 706 60790 or email publications\_graphics @fca.org.uk or write to Editorial and Digital Department, Financial Conduct Authority, 25 The North Colonnade, Canary Wharf, London E14 5HS.

## Contents

#### Summary

1 Overview

#### 2 Publication bias and replication

#### 3 A firm response: field trials and regulated firms

Experiment 1: Encouraging compliance: improving submissions to the FCA by mutual societies

Experiment 2: Help is here: helping firms apply for authorisation

#### 4 Towards action: consumer field trials

Experiment 3: Building a letter: engaging customers about their interest-only mortgages

Experiment 4: Please take your cash: encouraging consumers to claim redress from incomplete cash machine (ATM) transactions

Experiment 5: Just for you: using personalisation to attract attention

## 5 Under the microscope: laboratory experiments into consumer understanding and choices

Experiment 6: Invest or consume: testing the framing of retirement decisions Experiment 7: How much? Designing optimal price comparison websites in the payday lending market

Experiment 8: Shop and save: designing an annuity comparison tool

#### 6 Lessons learned

#### References

## Summary

The FCA has been at the forefront of the use of behavioural science and experiments to inform regulation. Since our first field trial on customer compensation in 2013, we have published the results of experimental research in a number of consumer markets, including savings accounts, structured savings products, and car and home insurance. This round-up paper presents eight further experiments, comprising five field trials and three online experiments, which test the effect of interventions that draw on behavioural theory, such as increasing salience or personalisation.

We investigate diverse questions including: How can we design disclosure about annuities to help people get a better deal? How can firms improve customers' engagement with their mortgages? What messages encourage customers to claim compensation? We also apply behavioural insights in a novel setting; improving compliance and engagement amongst regulated firms using communications.

While some experiments corroborate existing research or find interesting effects, others did not find any statistically significant effects. We are publishing these results, including non-significant and negative results, in the spirit of good research: improving evidence, combatting publication bias and making our research transparent. We also share some of the practical lessons we have learned, in the hope that others may benefit from them.

## **1** Overview

#### Background

The FCA published *Applying behavioural economics at the FCA* in 2013. An important recommendation of the paper was the use of field and laboratory experiments to diagnose problems and test remedies. Since then, we have published the results of experimental research in a number of consumer markets including cash savings (Adams, Hunt, Vale, & Zaliauskas, 2015; Adams, Hunt, Palmer, & Zaliauskas, 2016), structured products (Hunt, Stewart, & Zaliauskas, 2015) and general insurance (Adams, Baker, Hunt, Kelly, & Nava, 2015).

In this round-up paper, we present the results of eight experiments into communications and disclosure covering topics ranging from compliance to choosing a payday loan. This paper summarises a number of previously unpublished trials, which we conducted when we first began using experiments for regulation, as well as newer research conducted in collaboration with others. The research breaks into new areas for the FCA, including investigating interventions aimed at the firms we regulate, rather than consumers, as well as bolstering our evidence base on topics we have considered in previous Occasional Papers, such as customer compensation ('redress'). Many of the trials test similar treatments in different contexts, for example bullet points (salience), or interventions using personalisation. This allows us to see how interventions based on behavioural theory can lead to different responses in different contexts.

While some trials corroborated existing research or found interesting effects, for example that using the word 'annuity' reduces the number of people choosing an annuity, others did not find any statistically significant effects. In publishing these results, we are following in the footsteps of a number of major organisations, which report all their findings, including those that are non-significant, in order to combat publication bias, improve evidence and make research more transparent. We also share some of the important lessons on the practical experience of running experiments that we have learned along the way and report these in the final section of the paper. We aim to publish a separate paper on some of the broader issues of using experimental evidence to design policy in the near future.

#### Publication bias and replication

Publication bias, the selective publication of research which finds positive results, is a significant problem in social science research. It is caused by researchers and journal editors failing to publish experiments which might be less 'interesting', statistically non-significant or problematic for a prevailing theory and it results in published research that is systematically unrepresentative of the population of completed studies.

This can cause a knock-on problem for regulators and other policymakers who see the published papers, but do not see the studies which were carried out and then filed away, and so have to rely on a potentially distorted picture of the evidence. The problem is compounded by a 'replication crisis', the fact that many past experiments failed to produce the same results when replicated. In some cases, failure to replicate may be caused by the original researchers using samples of participants which are too small to detect effects reliably and consistently. This raises the risk of researchers using their discretion to report only positive results, or otherwise increase the possibility of finding statistically significant results.

4

Several leading research organisations, such as AllTrials, Cochrane and the Abdul Latif Jameel Poverty Action Lab, have taken steps to try to prevent publication bias. Measures include publishing registers of planned trials, so that statistically non-significant results cannot be ignored, or pledging to publish the results of all trials. Others, like the UK Behavioural Insights Team and the US Social and Behavioural Science Team, publish round-ups like this one, which present the results of all trials they have carried out. There is also an increasing focus within social science disciplines on replicating past results, including in new contexts. By publishing this paper, we hope to contribute to both these aims. We thus present our full set of trials to date, including those with results that are statistically non-significant, to allow others to learn from them and to enable replication. We also discuss this more fully in the second section, **Publication bias and replication**.

#### Experiments

This paper describes eight randomised controlled trials (RCTs). In a RCT, we randomly allocate participants to one or more groups and we treat one group as normal - this is the control group. We give the other group(s) our treatment and call it/them the treatment group(s). We then compare outcomes between the treatment and control groups.

RCTs can either be carried out in a real life environment, with real customers or firms making real decisions ('field trials') or by inviting participants into the 'laboratory' – either physically or online - to take part in the research ('lab experiments'). This paper describes five field trials and three lab experiments.

RCTs have the advantage of showing us the **causal** impact of our interventions. Field trials in particular may also allow us to understand the size of this impact and what difference it will make to consumers in the real world. For example, we can use a RCT to pilot a number of variations of content, tone, medium or style of a proposed communication and find out what really drives consumer behaviour and encourages action, to inform wider implementation.

Lab experiments can also inform the policy making process by creating a simplified model of the world in which to test specific decision making processes. The quantitative results of lab experiments may be less generalisable to the real world (this is referred to as 'external validity') compared with field trials, but a well-designed lab experiment should be able to give us qualitative results that we can use to generalise outside of the lab (see Kessler & Vesterlund, 2012, for a fuller discussion). There may also be practical reasons to use lab experiments instead of field trials, for example when evidence is needed more quickly, when field experiments are infeasible or unnecessary, or when we cannot find firms willing to partner with us to test interventions.

In this paper, we test a number of treatments, including some similar treatments across different experiments, for example, salience and personalisation. By setting each experiment in its broader context and re-testing interventions in different settings, it becomes possible to build a body of knowledge about behavioural theories (Deaton & Cartwright, 2016).

We describe the experiments in the three middle sections of the paper. The first section, A firm response: field trials and regulated firms, tests how to increase engagement amongst the firms that the FCA regulates. The second section, Towards action: consumer field trials, describes the results of three field trials which aimed to increase consumers' engagement with their financial products and nudge them towards action. The third section, Under the microscope: experiments into consumer understanding and choices, explores through a series of laboratory experiments how consumers use information to help them make decisions and how to frame information to help them make more rewarding choices. For summary statistics and regression analyses, please see the separate Technical appendix published alongside this paper.

5

A summary of the trials, treatments and results is below. Ticks ( $\sqrt{}$ ) denote that the treatment increased the desired behaviour and crosses (X) show that the treatment decreased the desired behaviour. Dashes (--) show that the treatment was not statistically significantly different from the control.<sup>1</sup>

Experiment	Context	Treatments	Result
1. Encouraging compliance: improving submissions to the FCA by mutual societies	Field trial on letters to help firms return their accounts data on time	Salience (bullet points)	
		Envelope warning	
		Different timings	$\checkmark$
<b>2. Help is here</b> : helping firms apply for authorisation	A/B testing of email subject lines to help firms with authorisation	Personalisation	$\checkmark$
		Salience (mentions FCA)	$\checkmark$
		Novelty	Х
<b>3. Building a letter</b> : engaging customers about their interest- only mortgages	Field trial testing letters to engage customers about their interest-only mortgages	Risk warning	Х
		Personalisation	
		Salience (bullet points)	
		Friendly style	$\checkmark$
<b>4. Please take your cash</b> : encouraging consumers to get redress from incomplete ATM transactions	Field trial testing letters to encourage customers to claim redress	Salience (bullet points)	( <mark>X</mark> )
		Salience (ease of process)	
5. Just for you: using	Field trial on the effect of	Personalisation	
personalisation to attract attention	handwritten envelopes on customers' likelihood to vote and claim redress		
<b>6. Invest or consume</b> : testing the framing of retirement decisions	Laboratory experiment testing the framing of retirement decisions	Framing	$\checkmark$
<b>7. How much?</b> Designing optimal price comparison websites in the payday lending market	Laboratory experiment testing potential regulatory standards for payday lending price comparison market	Decreasing effort	$\checkmark$
		Interactivity	
		Distraction: advertising	
8. Shop and save: designing an annuity comparison tool	Laboratory experiment into the effects of information prompts on shopping around for annuities	Personalisation	$\checkmark$
		Numerical examples	$\checkmark$
		Loss aversion	$\checkmark$

#### Sharing lessons learned

The FCA was the first financial regulator to use experiments to inform regulation. As such, we faced a steep learning curve in our endeavour to build practical experimental research into regulation. We have learnt a great deal from the experience of conducting these studies, particularly in relation to trial design, implementation and analysis. In the spirit of Dean Karlan

<sup>&</sup>lt;sup>1</sup> Statistically significant means that we would be likely to achieve these results in less than 1 in 20 cases if they were caused by chance (hence there is likely to be a real effect). This is not the same as economic significance which describes the importance of the effects to the issue in question (for example that an X% increase in responses equals a £Y increase in payments). It is possible to find effects that are statistically, but not economically significant, for example because the effect is very small.

and Jacob Appel's book *Failing in the field*, which details trials in which they made mistakes (Karlan & Appel, 2016), we include trials in this paper where our treatments did not have the intended effect, or where we recognise design flaws. In some cases, not finding an effect is inconsistent with previous research in similar but different contexts. We include these trials because they add to our overall understanding of how and when particular interventions do and do not work.

In the final section, **Lessons learned**, we explore a number of lessons on the practical experience of running experiments, covering themes such as:

- problem diagnosis,
- choosing the right method,
- pre-testing before going into the field,
- post-experiment analysis, and
- learning from and communicating null results.

We continue to learn as we carry out research, and through this scientific process, further questions may arise. As this paper is primarily a round-up of communications experiments which we are publishing for the purposes of transparency, its scope does not include broader discussion and recommendations in relation to integrating research into policy. However, there are many important questions in this area, such as how to select the most appropriate research methods to achieve different evidential standards and how we should interpret and apply the results of research to policy.

As noted, the field experiments in this paper deal primarily with optimising the FCA's or firms' communications rather than to develop remedies to implement as rules for the market. In those cases where we may use experimental evidence to inform policy, we need to take account of a number of issues, including policy timelines, the challenges of recruiting firms as voluntary partners and the strengths and limitations of different types of evidence. For example, while experimental evidence is one important source of information to guide our decision making, it should be viewed in its context and weighed up on its particular merits. We aim to discuss these wider issues, when it comes to interpreting and applying the results of research to policy, in a further publication dedicated to this topic.

7

## **2** Publication bias and replication

Like many organisations, the FCA is committed to evidence-based policy. For this to be effective, we rely on strong and robust evidence, which is representative of all relevant research that has been carried out and which can be replicated. Unfortunately, it is common for researchers and journal editors of all disciplines to fail to publish research which shows no significant effects (Schooler, 2011; Sterling, Rosenbaum, & Weinkam, 1995) or whose effects contradict previous research. This is called publication bias. It means that the body of research which has been published and peer-reviewed is no longer representative of all the research which has taken place and may be biased in favour of research which shows surprising or interesting results. When combined with a low level of replication for many previous experiments, this means that current beliefs informed by research evidence could be incorrect.

#### Combating publication bias

To put the effects of publication bias into context, imagine that 100 researchers carry out similar experiments and all are hoping to find an effect. This effect is conventionally measured by a result which would occur by chance less than five percent of the time. Five of them are likely to find an effect by chance using this measurement, regardless of whether one exists and these 5% have a strong incentive to attempt to publish their findings. The remaining 95 researchers do not find an effect (which is correct, as there is no effect). If the 95 researchers therefore decide that their research is uninteresting and unlikely to be chosen by a journal editor, their research may be consigned to a desk drawer. A new researcher to the field then reads all the published research (five studies) and concludes that 100% of experiments found this effect; it is hence, real.



In fact, examples like this are not so far from the truth. Ioannidis (2005) argues that most published research findings are false and in many fields simply reflect prevailing biases, due to low-powered studies and large numbers of researchers 'in a chase of statistical significance' among other reasons. Illustrating this, Franco, Malhotra and Simonovits (2014) analysed a known population of 221 social science studies and found that strong results – those where all or

most hypotheses were supported by the statistical tests - were 60 percentage points more likely to be written up and 40 percentage points more likely to be published than null results.

Publication bias, as illustrated above, means that researchers may not have access to the full body of evidence that exists and that they, and policymakers, may reach conclusions based on a distorted picture. This is particularly important when it comes to meta-analyses, studies which use statistical techniques to combine the results of multiple scientific studies to give a broader picture. While some studies use techniques such as funnel plots to try and identify and account for publication bias, its existence has the potential to distort conclusions. Further, publication bias is not the only way in which published research becomes unrepresentative of the total body of research carried out. Other forms of bias that could affect how we learn from past research (including meta-analysis) include: language bias (selective inclusion of studies published in English), availability bias (selective inclusion of studies that are easily accessible to the researcher) and familiarity bias (selective inclusion of studies from one's own discipline) (Rothstein, Sutton & Borenstein, 2005).

A range of organisations have committed to reducing publication bias, including 705 research organisations who have signed up to the AllTrials campaign. This campaign is calling for all past and present clinical trials to be registered and their full methods and summary results reported, so that researchers, patients and policymakers are able to make accurate and informed decisions. Other networks and organisations active in this area include Cochrane, which produces systematic reviews of primary research in medicine; the Campbell Collaboration which does the same for research in crime and justice, education, international development and social welfare; and, the Abdul Latif Jameel Poverty Action Lab, which maintains a searchable database of 774 randomised evaluations in 69 countries including many on-going experiments, as well as a large range of resources for researchers and policymakers to create evidence-based policy.

The FCA supports the aim of reducing publication bias wherever possible and we are therefore publishing the full range of our experiments to date, including those where we found no effect of our interventions. We also see value in sharing our experiences with others, as we do with the practical lessons learned in the final section of this paper, and plan to publish further lessons learned in relation to the theoretical considerations of carrying out research for policy.

#### Enabling replication

Interventions that are highly effective in one context may not work in another, which is one reason why testing results in different contexts is so important. In fact, we should not expect replication across different contexts and across different study populations. Failure to replicate does not make either finding useless – "we can often learn much from coming to understand why replication failed" (Deaton & Cartwright, 2016).

Attempting to replicate specific experiments (rather than testing similar interventions in different contexts) is also crucial to developing theories, through confirming or disconfirming existing results and by helping us understand the size and limitations of effects (Brandt et al., 2014). However, many academics are concerned that there may be a replication crisis in social science, with one study replicating successfully only 39% of effects in 100 articles from high-ranking psychology journals, with the remainder either inconclusive or failing to replicate (Open Science Collaboration, 2015).

One reason for the low levels of replication is that many experiments use sample sizes which are too small to be able to detect effects with any confidence. When experimenting, researchers try to reject the "null hypothesis" - that the treatment has no effect. They will need enough participants to ensure that the experiment has a high probability of being able to do this (this is called "statistical power"). However, many experimenters fail to recruit enough participants to reach a sufficient level of power (normally 0.80 for social science experiments). In fact, low powered

9

experiments are a long-standing problem, from Cohen's (1962) study showing a median power of 0.46 for medium size effects in the Journal of Abnormal Psychology to a replication in 1989 which (unusually!) showed similar results (SedImeier & Gigerenzer, 1989). One explanation for the prevalence of low powered studies is the strong intuitions that both naive subjects and trained scientists have about random sampling and small numbers (Tversky & Kahneman, 1971).

When insufficiently powered experiments are paired with a bias towards publishing results which are interesting, surprising or consistent with a researcher's previously published work, it raises the possibility that many published studies simply show results that come about by chance rather than real effects. In fact, researchers have many degrees of freedom when running experiments, for example, making subjective decisions on when to stop data collection, how to deal with outliers and which outcome measures to report (Simmons, Nelson & Simonsohn, 2011; Mathieu, Boutron, Moher, Altman & Ravaud, 2009). They may compensate for small sample sizes by using practices which increase the probability of finding, by chance, significant effects that are not real.

In this context, we would like to be open both about our research and our experience of carrying out research for policy, in order to help other regulators, public bodies and firms who are grappling with similar questions and to enable replication.

## **3** A firm response: field trials and regulated firms

There is a large and growing body of research on consumer behaviour, looking at how consumers respond to and interact with communications from firms and products, but relatively little research which examines communications between regulators and the firms they regulate.

In the following section, we present two field trials aimed to help us understand how to increase engagement with firms and help them to understand and follow our rules.

## Experiment 1: Encouraging compliance

Improving submissions to the FCA by mutual societies

Researchers: Paul Adams, Robert Baker, Helena Robertson and Laura Smart

#### Objective

To help mutual societies (a type of firm that must be registered by the FCA) submit their annual returns to the FCA on time.

#### Background

The FCA receives the annual returns and financial accounts of mutual societies in our capacity as registering authority for these societies (distinct from our role as a financial regulator). Societies must submit their annual accounts at specified points in the year based on their financial year-end date. This was historically defaulted to 31 December, but societies can choose to change this if they wish. In the past, not all societies have submitted their return on time and some failed to submit at all. Late submissions or failure to submit costs the FCA time and resources to chase societies. It also has consequences for the societies who can lose their registration or receive a fine after successful prosecution by the FCA.

There are many possible reasons why mutual societies fail to provide annual returns on time. Societies are often small organisations run by a small number of volunteers. They may exhibit inertia, and societies who have never submitted annual returns have little impetus to start doing so now (in our sample, 8.55% had never submitted returns and 7.73% had not submitted returns for three years prior). Other possible explanations include forgetfulness, given that the time allowed for societies to submit after the year-end is a period of seven months or a change of society secretary - the new secretary may not be made aware of the requirement to submit the annual return and accounts by the outgoing secretary.

We wanted to see if there were specific messages we could use when communicating with societies, which would draw their attention and encourage them to submit on time.

#### Method

We sent letters to a sample of 7,984 mutual societies<sup>2</sup> who were stratified by type of organisation, the month of their financial year-end and the last year they took action, and then randomly allocated to different treatments or to a control group. The treatments were:

- **Bullets**: Including salient bullet points and a message about penalties: "Last year mutual societies like yours were fined up to £3,000 for failing to provide this information on time";
- **Warning**: Adding a warning to the envelope: "IT IS A LEGAL OBLIGATION TO COMPLETE AND RETURN THE ENCLOSED FORM"; and
- *Timing*: Sending the letters on different dates (26 May, 3 June or 8 July). This helped us to estimate the effects of the length of time between the letter and the deadline on compliance.

The treatments were designed to test some of the behavioural insights that have been effective in other contexts, including reminding recipients of penalties (Fellner, Sausgruber, & Traxler, 2013) and legal requirements (Koelle, Lane, Nosenzo, & Starmer, 2016).

We also combined treatments to create 12 different groups; for example, one group was sent a letter on 26 May with no bullet points and no envelope warning and another was sent a letter on 3 June with bullets but no envelope warning. Please see Tables 2 and 3 in the Technical Appendix published alongside this paper for details of the balance of different demographics in each of the groups.

For the purpose of testing the effects of bullet points and warnings, the control group was formed of those who received no bullet points and envelope warning, but whose letters were sent on all three of the above dates. For the purpose of testing the effects of timing, the control group was the first cohort (26 May), but all groups contained a similar number of people who received bullet points, envelopes and both or neither.

We measured whether societies had interacted with the FCA in two ways ten months after the communication. Either the society had submitted its annual return to the FCA or the society's status had been changed towards losing its status as a mutual society (an indication that the society was closing, either initiated by the society itself or the FCA).

#### Results

Of the societies in the trial, 6,456 took action (80.9%), while 1,528 (19.1%) did not take any action.

Figure 1 shows the effects of bullets points and envelope warnings on all letters sent (i.e. combining letters sent on different dates). Bullets and warning treatments failed to change the behaviour of societies compared to the control.

<sup>&</sup>lt;sup>2</sup> Specifically, industrial and provident societies





Figure 2 below shows the effect of timing of the letter on rates of action on all letters sent (i.e. combining control, bullets, envelope and bullets and envelope letters), but excluding societies who were not randomly assigned to timing treatments because their deadline was too early<sup>3</sup>. The timing of the letters did affect the firms' response. Societies who were sent a letter in July (and who therefore had, on average, a shorter deadline) were 2.4 percentage points more likely to respond to the communication. Across all groups, those with a shorter deadline were more likely to respond than those with a longer deadline. Please see regression outputs in Tables 4 and 5 in the Technical Appendix for more details.



Figure 2: Action rates by timing treatment

We also found a relationship between likelihood to take action and whether or not societies chose their own end of year accounting date or stayed with the default. Societies who chose their own accounting end date were less likely to take action than those who kept the default date of 31 December (and hence a submission deadline of 31 July), though it is difficult to separate this from possible seasonal effects. Please see Table 6 in the Technical Appendix for details.

<sup>&</sup>lt;sup>3</sup> Some firms were allocated to the May or June groups because their deadline for submitting their return was before the date of the June and/or July letters. We include regressions which include and exclude these firms. Results presented in the letters graph include these firms and results in the timing graph exclude these firms.

#### Conclusions

The most likely reason why letters sent in July led to higher rates of action is that those in July had a shorter deadline than other groups on average (median of 23 days for July rather than 66 for May and 58 for June). This is consistent with the negative correlation we see for all treatment and control groups between the time until deadline and action.

It is possible that the timing of the letters in this case was more important than the content. This is similar to the finding in Adams, Hunt, Vale and Zaliauskas (2015), that the fact of a reminder relating to a decrease in customers' savings interest rates was more important than its precise phrasing. Simply sending a reminder increased switching by at least 8% relative to not receiving one.

The relationship between accounting period (and hence deadline) choice and likelihood to take action is consistent with Ariely and Wertenbroch (2002), who found that students imposing their own deadlines for submitting papers performed more poorly than those who were given default deadlines that were evenly spaced. This was contrary to our initial expectations, where we expected those who had changed away from the default deadline to be more engaged and therefore perform better, again proving the importance of testing to dispel our own preconceptions.

While in this case, resource constraints limited the amount of prior investigatory work we could do, it may be useful in cases like this to consider qualitative work to give a greater understanding of the reasons firms might not submit on time, which could be used to tailor treatments to specific findings.

## Experiment 2: Help is here

Helping firms apply for authorisation

Researchers: Laura Smart, Dom Suckling and Roisin Wilson

#### Objective

To increase engagement with FCA emails of firms applying for authorisation

#### Background

In April 2014, the FCA took over responsibility for consumer credit from the Office of Fair Trading (OFT). This meant that nearly 50,000 firms previously licensed by the OFT were required to apply for FCA authorisation if they wanted to continue with their consumer credit activities. Between 1 April 2014 and 31 March 2016, we received nearly 37,000 applications for authorisation from consumer credit firms

Each firm was allocated a three-month window in which to apply and the FCA sent out a series of emails both before and during this to help firms with their application. One difficulty with providing information by email is engagement; many firms simply did not open the emails and so could not benefit from the information. We carried out a series of simple randomised trials aimed to increase engagement with one of the emails that gave firms a link to videos where they could learn more about the application process.

#### Method

For each email sent, we divided the recipients into two random groups using email distribution software. Each group received a different email subject line designed to attract attention and overcome inertia. We then measured email open rates for each group. We repeated this over four cohorts of different firms in iterative A/B tests (i.e. the best performing message is then pitted against a new message). A/B testing is the simplest form of randomised controlled trial, pitting one control against one alternative and measuring the outcome. It can be a quick and inexpensive way to test and adapt communications over time, and is especially used in digital media such as websites and emails.

In our A/B tests, we used the following messages:

- Your FCA application: help is here to complete your section;
- Watch our new step-by-step guide videos: helping you apply for authorisation;
- Video: new FCA guides to help you complete your application;
- New: step-by-step video guides to help you complete your application; and
- [Firm Name] help is here to complete your application.

We chose messages for which there was anecdotal evidence from practitioners in previous contexts, as well as messages which relied on behavioural evidence such as personalisation ('[Firm name], help is here...'), salience ('Your FCA application...') and novelty ('Watch our new...'; 'Video: new FCA guides...'; 'New: step by step...') (Wu & Huberman, 2008; Behavioural Insights Team, 2014).

We measured email open rates for each subject line and take that as an indicator of engagement with the material.

#### Results

A summary of the results is set out in Figure 3 below. Trials 1 and 3 produced statistically significant results, showing that the email subject '*Your FCA application: help is here to complete your section*' led to more firms opening the message than '*Watch our new step-by-step guide videos: helping you apply for authorisation*' or '*Video: new FCA guides to help you complete your application*'.

Many of the trials showing non-significant results are low-powered, due to constraints on cohort sizes. Due to the relative homogeneity across the cohorts, where possible we repeated experiments with the same subject lines (trials 2 and 3, 5 and 6) in order to increase the sample size. Trial 3 produced a statistically significant result both alone and combined with trial 2, while combining trials 5 and 6 did not result in significance. Please see Table 7 in the Technical Appendix for further details.





#### Conclusions

The most successful subject line was the original 'Your FCA application: help is here to complete your application', along with '[Firm name] help is here to complete your application'. It is possible that the mention of the FCA (salience) and personalisation in the above messages were effective in encouraging engagement, though this is difficult to measure precisely as the subject lines varied on more than one attribute. In future experiments, it may be worth varying subject lines more systematically, so that we are able to separate any effects that we see and attribute them to specific behavioural theories. For example, we could test the effects of personalisation by pitting 'Help is here...' against '[Firm name]: help is here...' or the effects of reciprocity by testing 'We've made you a video; please click here' against 'Video: please click here'.

Whilst we found limited effects here, the use of A/B testing can be a cheap and effective way for organisations to improve their communications, especially where technology already exists to vary information, where outcomes are already measured routinely and where an iterative approach can make marginal small improvements.

## **4** Towards action: consumer field trials

Our previous research on compensation communications, as well as broader letter-based communications in the general insurance and cash savings market found that small changes such as simplifying key messages or using loss aversion increased rates of action (Adams & Hunt, 2013; Adams et al., 2015; Hunt et al., 2015). The following field trials test similar small changes, many based on behavioural economic theory, with some surprising results. These changes include **salience**, drawing attention through putting main messages front and centre, **simplicity**, reducing cognitive effort by making information easy to understand and **personalisation**, attracting attention and increasing favourability through greetings and personal messages.

## Experiment 3: Building a letter

Engaging customers about their interest-only mortgages

Researchers: Paul Adams and Will Brambley

#### Objective

Encourage consumers to consider how best to repay their interest-only mortgage.

#### Background

In 2013, over two and a half million customers had interest-only mortgages (i.e. where the loan is not repaid until the end of the mortgage term). Some 600,000 of these mortgages were due to mature before 2020. FCA research showed a significant minority of people did not have repayment plans in place to repay their loans at the end of their term. Therefore, in May 2013, the FCA published guidance on dealing fairly with interest-only mortgage customers who risk being unable to repay their loan.<sup>4</sup>

The FCA and the industry wanted to increase engagement from customers with the letters that firms were providing. This was to ensure that customers thought about their repayment plans and to encourage them to contact their providers to discuss potential options. We worked with one provider to test the specific messages included in the letters they sent out over a six-month period in 2013.

We know from previous research carried out by the FCA and elsewhere that the specific wording and framing of messages in letters can have profound effects on consumer engagement and subsequent actions (Adams & Hunt, 2013). The current research had two aims – to understand how results from other research might be generalisable to this context; and, for this specific case, to help inform interest-only mortgage providers, the Council of Mortgage Lenders and the Building Societies Association.

<sup>&</sup>lt;sup>4</sup> http://www.fca.org.uk/news/interest-only-mortgages

#### Method

We worked with an interest-only mortgage provider to run a randomised controlled trial. Customers (n=8,004) were stratified by loan to value ratio, age, time to the end of the mortgage period and remaining mortgage outstanding and then randomly assigned to the control or one of four treatment groups.

The five letters were:

- 1. Control: Standard letter written by the firm;
- 2. *Riskless:* Removing the standard repossession risk warning ("Your home may be repossessed if you do not keep up repayments on your mortgage");
- 3. *Non-personal*: Removing a table of personal data which included balance and time left on mortgage;
- 4. Bullets: Summarising the key information in bolded bullet points at the top of the letter; and
- 5. *Friendly*: Removing the risk warning and re-writing the letter to be friendlier and more informal in tone.

Please see Table 9 in the Technical Appendix for details of the balance of different demographics in each of the groups.

The firm was able to monitor a number of outcomes including:

- Whether the customer proactively contacted the provider; and
- Whether the customer was open to discussing potential repayment options when contacted by the provider.

#### Results

After removing duplicates and those in arrears or who had complained, who may have been treated differently, we analysed data from 7,319 customers.

The rate of response was low (5.5% in the control group). Customers were more likely to respond if they had opened the account within the last six years, were within two years of maturity or if they received the letter in July or October.

Removing the repossession risk warning and making the letter more informal (Letter 5) led to a 2% increase in responses, as shown in Figure 4 below. Adding bullets (Letter 4) reduced the response rate by 1.8% and removing personal information made no difference from the control.





#### Conclusions

The results of this trial show that less is more in this context and simplicity is likely to improve response rates. It appears that removing the risk warning on this letter (the warning was not mandated in this case) actually increased the response rate, particularly when combined with a more informal tone. It is possible that this is because the risk warning might scare customers and lead them to put their head in the sand rather than engage with the firm. However, this result should be seen in the context of this experiment rather than extrapolated, as for example, we are not able to observe the potential long-term educational benefits of the risk warning nor of the effects of risk warnings in other types of product information, such as advertisements.

It is also notable that bullet points reduced response rates, in contrast to the findings of Adams & Hunt (2013). This is likely to be due to contextual factors; for example, the situation and product in this trial was more complex and the call to action may have been less predictable (calling to discuss plans for a mortgage product, as opposed to claiming redress).

## Experiment 4: Please take your cash Encouraging consumers to claim redress from incomplete cash machine (ATM) transactions

Researchers: Paul Adams and Robert Baker

#### Objective

Encourage consumers to reclaim money they lost as a result of not taking bank notes requested from and presented by the ATM.

#### Background

Customers of a firm had attempted to withdraw money from ATMs between 2004 and 2011 but had failed to take the money from the machine when it was offered. The money was then debited from their account as normal. Consequently, a redress exercise was undertaken in 2012/2013 to return this money. Some of the customers entitled to redress had since closed their accounts and therefore the firm needed to write to them to get new bank details in order to repay them.

The average (mean) redress due was £95, which included the original value that had been debited but not withdrawn from the ATM and an additional interest payment dependent on the length of time since the failed transaction.

#### Method

We worked with the firm to run a randomised controlled trial to test whether different ways to simplify or frame the information might improve customer take-up of redress. We tested the firm's original design with two changes: using bullet points and adding information about the process for claiming. We tested four letters:

- 1. Control: Standard letter written by firm;
- 2. **Bullets**: Summarising key information in bullet points at the top of the letter, including the amount that was owed and why and the action the customer could take, with the message "ACT NOW: You are less likely to respond if you delay";
- 3. **Process**: Summarising the ease of claiming redress in a numbered list in the body of the letter ('1. Complete the attached form within 28 days, being careful to fill out the correct account details. 2. Post the form to us using the provided pre-paid envelope'); and
- 4. Bullets and process: Both treatments included.

A total of 5,589 customers were randomly allocated to receive one of the letters, using a stratified sampling approach which took account of all observable characteristics: the length of time since the ATM use, the amount the customer tried to withdraw, gender and one further characteristic which we redact to protect the anonymity of our partner. As some customers received letters with both changes, we were also able to investigate whether there were any interaction effects. Please see Table 12 in the Technical Appendix for details of the balance of different demographics in each of the groups.

The firm was able to monitor a number of outcomes including:

- Whether the customer returned the form and therefore claimed the redress; and
- Whether the customer contacted the firm and for what reason.

#### Results

Around half of the customers received redress following the trial. In total 2,839 (50.8%) consumers made a claim and received an average redress amount of £95. As expected, those owed more were more likely to make a claim. We also find that more recent transactions were more likely to be reclaimed, which is likely to be a result of reducing address accuracy over time. In fact, this means that the proportion of eligible customers claiming redress could be much higher than we observe. In line with previous research from the FCA we find women are more likely to respond then men. Joint account holders were even more likely to respond, perhaps because these individuals are more likely to have stable addresses.

We find that the *Bullets* and *Process* treatments do not affect whether customers claim redress, as shown in Figure 5 below. This is in contrast to the findings of FCA Occasional Paper 2 (Adams & Hunt, 2013), which found that both salient bullets and text explaining the process for claiming increased the proportion of consumers claiming redress. We note that in this instance the average amount claimed is much higher (£95 in this research compared to £21 in Adams & Hunt (2013)) and the proportion of customers claiming redress in the control group is also much higher - in this instance over 50% of customers claimed redress.



While not affecting the level of redress, the bullet treatment led to significantly more customers contacting the firm, as shown in Figure 6 below. Many calls were regarding the letter's authenticity and concerns that the letter might be a phishing scam. Like all the letters, the customer was asked to provide bank details, but perhaps the direct way in which the *Bullets* letter phrased this request made some people uneasy.



Figure 6: Contact by treatment

Please see Table 13 in the Technical Appendix for regression outputs.

#### Conclusions

The trial shows us that context is very important when designing interventions. While efforts to make communications simpler, such as bullet points and summarising the process to claim might work in some redress cases, in others there may be a risk that changes in directness or style cause customers to question the authenticity of the letter. It is important to be mindful of the specific context of a communications exercise to avoid unintended effects. Where possible, qualitatively road-testing treatment ideas - for example, through user experience testing or interviews - may unearth unexpected effects before putting them into the field.

## Experiment 5: Just for you

Using personalisation to attract attention

Researcher: Laura Smart

#### Objective

To encourage customers to vote on a scheme to review the sale of insurance products.

#### Background

We worked with a firm which was writing to policyholders of an insurance product as part of a redress exercise. The first part of the exercise involved giving customers the opportunity to vote on a proposed scheme for the company to review the way the product was sold.

In many cases, responses to redress exercises and other letter-based communication campaigns can be low (Adams & Hunt, 2013). There is some evidence that handwritten communications can increase salience and encourage more recipients to open letters and take action (Irish Revenue, 2013). We wanted to find out whether handwritten envelopes would increase the number of customers who responded in this case.

#### Method

We were restricted to selecting a quasi-random sample of 100 customers from one cohort of those receiving letters (n=20,288) to receive handwritten envelopes. For logistical reasons, the 100 selected were the first consecutive group of 100 customers from the customer list who had not already been in contact with the firm about their insurance product. After removing duplicates (288) and those who had already contacted the scheme (648), the total sample used in the experiment was 19,352, with a control group of 19,252 and a treatment group of 100. The treatment group of 100 customers received letters in handwritten envelopes and the remaining customers formed a control group, receiving letters in printed envelopes. Please see Table 15 in the Technical Appendix for details of the balance of different demographics in each of the groups.

The firm was able to measure:

- Whether the customer voted on the scheme; and
- Whether the customer claimed redress and whether their claim was upheld.

#### Results

In the whole sample (n=19,352), 4,071 customers (21%) voted and 5,462 (28%) got in contact with the firm in relation to redress, with 5,411 customers (28%) ultimately receiving redress.

Analysis showed that randomisation was effective and customers in the treatment and control groups were similar on observable characteristics, including age, gender and region (see Table 15 in the Technical Appendix for details).

There were no statistically significant effects of handwritten envelopes on voting rates or rate of claiming redress, as shown in Figure 7 below.



Figure 7: Voting and claim rates by treatment

Please see Table 16 in the Technical Appendix for regression outputs.

We also looked at the effect of demographic characteristics on voting rates. Customers living in the North East were more likely to vote. There was an effect of age amongst those whose age we knew (about half the sample): both younger people and older people were less likely to vote, with the peak voting age around 65 years. Customers whose age was unknown were less likely to vote than those whose age was on file.

#### Conclusions

It is likely that the non-significant effect of the handwritten envelopes is because the trial was underpowered, due to logistical constraints which required a small treatment group. Analysis shows that the experiment had a power of 24.6% for voting rate and 9.6% for claim rates, which is much lower than the 80% usually recommended for social science experiments.<sup>5</sup> While the size of the treatment group was known before the start of the experiment, we had limited control over this and the trial itself was attractive in being relatively simple, since the redress exercise was going ahead anyway. However, it is important in such cases to balance the benefits of testing with the limitations of small samples and to prioritise potential trials accordingly.

<sup>&</sup>lt;sup>5</sup> We used the following post-hoc power calculator: <u>http://clincalc.com/stats/Power.aspx</u>. Power reported is based on the number of subjects in each group, the proportions voting or claiming and alpha at 0.05.

## **5** Under the microscope: laboratory experiments into consumer understanding and choices

Sometimes we need to be able to estimate how consumers might understand and interact with information in order to design effective policies. One method to do this, particularly in cases with little relevant existing evidence and when there are a large number of ideas to test, is the lab experiment.

In a lab experiment, we recruit participants according to certain criteria (for example: age or experience with the product in question) and invite them to participate in a short experiment, either in person or online for which they may receive a small financial reward. Lab experiments are usually randomised controlled trials and may test comprehension, memory or choices. They are particularly suitable when we want to find out why consumers behave in the way they do, as they allow us to test multiple treatments varying in small ways, and in some cases, to ask consumers directly about their reasons for their choices. Importantly, they give researchers close control over the whole decision making environment.

One limitation of lab experiments is external validity; will consumers behave the same way in the real world as they did in the lab, when they were aware they were taking part in an experiment? Where possible and appropriate, lab experiments can be supplemented by other research, including field trials.

For a more detailed analysis of the value of and suitable circumstances for laboratory experiments, please see FCA Occasional Paper 3 (Iscenko, Duke, Huck & Wallace, 2014).

In the following experiments, we see quite large effects of our interventions compared with the field trials in sections 3 and 4. There could be a number of reasons for this:

- Participants were recruited and incentivised to take part in the research and so may have been more motivated to pay attention and make good choices. However, we would expect that participants in the real world would also want to maximise their choices (notwithstanding constraints on cognitive load and time).
- Two of the experiments test differences in the choice architecture of existing mechanisms for providing information, for example price comparison websites and pension packs. This contrasts with information received unexpectedly, such as the letters or emails in the experiments above.
- Using lab experiments gave us more control over the design of the treatments and allowed us to test more radical (and therefore potentially more effective) changes, such as removing advertising, which it might be harder to test in the real world.

## Experiment 6: Invest or consume?<sup>6</sup>

#### Testing the framing of retirement decisions

Researcher: Robert Baker, Matteo Aquilina Delivered by YouGov

#### Objective

To understand how the framing of annuities can affect the decisions made by consumers.

#### Background

In the past many UK consumers were effectively required to buy an annuity (guaranteed income for life) with a large proportion of their pension pot. Since April 2015, UK consumers have had greater choice at retirement over how to access their pension savings.

In principle, how the income streams associated with different options are presented should not affect the consumer's decision – the actual income stream is unaffected by the wording. However, in practice we know that the framing of options does affect consumer decisions, and in predictable ways. In this experiment, we wanted to find out how consumer choices would change depending on whether we framed options in terms of **consumption** (how much money they could spend) or **investment** (what the return on their investment would be), as well as the effect of using the word 'annuity'.

#### Method

We presented a representative sample of 907 UK consumers aged between 55 and 75 with hypothetical choices between retirement income strategies, such as **buying an annuity** (a guaranteed income for life), or alternatives like **self-annuitising** (putting the money in a bank account which paid a set amount of interest for life) and **self-amortising** (consuming the money until it runs out). Participants made choices between different pairs of options five times during the experiment.

Participants were split into three random groups:

- **Consumption** group: information framed in terms of the amount of money they could spend in retirement and strategies were unlabelled,
- **Annuity consumption** group: information the same as the consumption group but the annuity option was named as an annuity,
- **Investment** group: information about the total size of the pot, the options were framed in terms of investment and returns and strategies were unlabelled.

#### Results

Those who received information about the options framed in terms of consumption preferred the annuity compared with alternative strategies and the opposite effect was seen for the investment group. For example, consumers were asked to choose between an annuity and a savings account from which they can draw the same income as the annuity, or alternatively they could only spend the interest on the account (therefore keeping their capital intact). In the consumption

<sup>&</sup>lt;sup>6</sup> For a more detailed report on this experiment, please see Financial Conduct Authority. (2014a) Does the framing of retirement income options matter?: a behavioural experiment.

frame, 66% of consumers chose the annuity. In contrast, under the investment frame, only 17% of consumers chose the annuity (see Figure 8).



#### Figure 8: Participant choices

Use of the term 'annuity' also affected how many consumers chose this as an option. For example, without using the term 'annuity', 66% of consumers preferred the cash-flows provided by the annuity to those provided by the savings account. In contrast, merely including the word 'annuity' in the choice reduces consumers' preference for the annuity to 50% - a 16 percentage point reduction. For the other two questions where consumers were asked to compare an annuity with alternative drawdown strategies, we could see a smaller impact of the term.

We also found that across the different frames, consumers without children had a greater preference for annuities compared to those with children, perhaps because they attach less value to the strategies which provide a bequest.

#### Conclusions

Consumers' retirement choices appear to be materially affected by the way the relevant information is presented to them. These finding are consistent with the results from Brown, King, Mullainathan and Wrobel (2008) who ran a similar experiment on consumers in the US. Framing annuity choice as an investment appears to bias consumers against annuities. Currently, consumers are provided with the value of their pot continually during accrual and at the point of retirement. This may create an investment frame through which consumers view annuities. While there are very good reasons why consumers should be given this information, it does appear to lead to consumer aversion to annuities. Such a finding means that it is vital that consumers get the right information and help at the time of retirement to ensure they make the best decision they can about their retirement income.

Further, consumers seem to associate the term 'annuity' with poor value products. The use of the term 'annuity' reduces the number of consumers who choose the annuity, even though consumers appear to value the underlying characteristics of the annuity.

## Experiment 7: How much?

#### Designing optimal price comparison websites in the payday lending market<sup>7</sup>

Researchers: Alessandro Nava and Qamar Zaman Delivered by YouGov and London Economics

#### Objective

To find out how the presentation of information on price comparison websites affects consumer choices.

#### Background

In 2015, the Competition and Markets Authority (CMA) discussed the potential for price competition between payday lenders in certain cases. The CMA's market remedies included an Order on all online payday lenders that will prohibit them from lending unless they publish loan product details on at least one price comparison website. Alongside this Order, the CMA made recommendations to the FCA to raise the standards of price comparison websites that compare payday loans to help to ensure better outcomes for consumers. The objective of this study was to test potential standards for price comparison websites that compare payday loans and inform the FCA's response to the CMA's recommendations.

#### Method

We recruited 808<sup>8</sup> UK consumers who had taken out a payday loan in the past 12 months and/or intended to take one out in the coming 12 months to take part in an online experiment followed by a survey. Participants saw three versions of a price comparison website listing hypothetical payday loans and in each case had to select the cheapest loan (defined as the one with the lowest total amount payable (TAP)) amongst the ones that best suited their needs in terms of amount and duration. All participants saw Treatment A, which encompassed all of the standards the FCA wanted to test, and were randomly allocated to see two of six alternative versions, each of which was missing one of the standards.

The treatments were:

- A. *All standards*: All loans on one page, ordered by TAP ascending, no banner advertising, input functionality and market coverage disclosure (see the Technical Appendix);
- B. *Featured loans*: As A, but showing only featured loans on the first page requiring a click-through to see all loans;
- C. No order: As A, but loans not ordered by TAP;
- D. *RAPR:* As A, but loans sorted randomly and displaying representative APR (RAPR) instead of TAP;
- E. Advertising: As A, but with banner advertising;
- F. **No input functionality:** As A, but no functionality to filter loans by loan amount and term; and
- G. No standards: none of the standards.

<sup>&</sup>lt;sup>7</sup> For a more detailed report on this experiment, please see London Economics and YouGov. (2015). High-cost short-term credit price comparison websites: a behavioural study for the Financial Conduct Authority.

<sup>&</sup>lt;sup>8</sup> A number of individuals had to be removed from the final data set for reasons of quality control and data robustness. Therefore, 791 respondents are included in the final analysis.

#### **Results**

When viewing the control page which contained all of the proposed standards (Treatment A), 63% of consumers chose the cheapest loan consistent with their preferences, as shown in Figure 9 below.

The largest single effect was observed when customers were required to click through to see all, not just featured loans (Treatment B). In this case, three percent of respondents seeing price comparison websites with loan offers across two pages chose the cheapest deal, compared to 63% of respondents seeing price comparison websites with all loan offers displayed on a single page.

The second largest single effect was observed when displaying RAPR instead of TAP (Treatment D). Respondents who were not explicitly shown the TAP chose the cheapest deal 13.5% of the time, compared to 63% of respondents who were shown the TAP.

The third largest single effect was observed when loans were sorted randomly by TAP, but with one restriction: the cheapest deals were never on top (Treatment C). Respondents who saw loan offers ranked randomly chose the cheapest deal 27% of the time, compared to respondents who saw loan offers ranked in ascending order of TAP who chose the cheapest deal 63% of the time.

We found no significant effects of the other treatments. Self-reported differences in payday loan usage (for example, whether participants had taken out a payday loan before) showed no significant effects on whether participants chose the best deal.



Figure 9: Participants' choice of cheapest loan by treatment

#### Conclusions

The research tested a number of specific interventions to improve consumer performance in finding the cheapest loan consistent with their preferences. The findings showed that requiring customers to click through to a second page to see all the loans significantly reduced performance, as did not providing the TAP. This is consistent with behavioural evidence that people often make poorer choices because of relatively small barriers such as needing to click-through, or larger barriers such as having to make a calculation (Behavioural Insights Team, 2014). It also shows that small changes to the way information is presented can have a large effect on behaviour.

## Experiment 8: Shop and save

#### Designing an annuity comparison tool<sup>9</sup>

Researchers: Paul Adams, Sumedha Pathak, James Ridgewell and James Shafe Delivered by Oxera and the Nuffield Centre for Experimental and Social Sciences

#### Objective

To find out whether and which information prompts might encourage customers to shop around for their annuity.

#### Background

The retirement income market study by the FCA found that competition is not working well for consumers in the annuities market (Financial Conduct Authority, 2016). In particular, the FCA found that, while most consumers who purchase their annuity from their current pension provider (80%) can benefit by shopping around, many fail to do so (Optimisa Research, 2013).

Consumers' decision not to shop around may in part be related to the way they perceive the associated costs and benefits. Many consumers say that shopping around is not worth it, despite clear evidence that they could increase their annuity income by switching away from their pension provider (Financial Conduct Authority, 2014b). Furthermore, they may believe that it would be difficult to evaluate the array of potential options available, and that their easiest and most straightforward choice is to stick with their existing provider. Moreover, some consumers may wish to avoid decisions they will regret in the future, and so end up staying with their current pension provider.

This experiment focuses on whether information prompts, given as part of the pension annuity quote from their pension provider, are likely to encourage more consumers to overcome inertia and shop around. A laboratory experiment was suitable for investigating these questions as we were interested in the purchase stage, where it is likely that consumers are paying some attention.

#### Method

We recruited a sample of 1,996 UK consumers aged 55-65 and allocated them randomly into six groups. Five of the six groups received an information prompt with varying information intended to encourage consumers to shop around. The sixth group were not presented with an information prompt, and therefore acted as a control against which the behaviour of the other groups could be compared. Treatments included:

- Personalised annual: participants were provided with the highest quote they could obtain by shopping around;
- Personalised lifetime: as above but with an estimate of the foregone gains from not shopping around over a typical person's lifetime;
- Non-personalised annual: participants were provided with an estimate of how much they could obtain by shopping around (not exact);
- **Non-personalised lifetime**: as above but with an estimate of the foregone gains from not shopping around over a typical person's lifetime; and

<sup>&</sup>lt;sup>9</sup> For a more detailed report on this experiment, please see Oxera and the Nuffield Centre for Experimental Social Sciences. (2016). Increasing consumer engagement in the annuities market: can prompts raise shopping around?

• **Call to action**: participants were told that "80% of people who fail to switch from their pension provider lose out by not doing so", with a visual representation.

Participants were incentivised to maximise their hypothetical retirement income through a series of tasks including choosing an annuity. "Shopping around" involved some additional effort (giving personal information and choosing a quote), as it would in real life. The experiment also aimed to induce status quo bias, inattention and fatigue by getting the participant to multi-task and answer a series of questions before making a choice.

#### Results

All the information treatments had a significant impact on the consumers' decision to shop around. Figure 10 below shows the proportion of participants who clicked to shop around for the six groups. The effects of the treatment, measured as the difference between the treatment and the control, range from around 8 percentage points for the non-personalised lifetime treatment to 27 percentage points for the personalised annual treatment.

Furthermore, there is substantial variation in the effects across the five treatments, with the personalised annual treatment achieving the highest impact on shopping around, followed by the call to action treatment. The two treatments prompt consumers in different ways: the personalised treatments offer information that is reliable and customised to the consumer; and the call to action treatment offers simple, easily digestible information accompanied by a strong social-comparison visual.

Moreover, the non-personalised treatments caused less shopping around than both the personalised and the call to action treatments, although this was not significant. There are two main potential explanations for this. The non-personalised treatments may have contained too much text, which may have led to information overload and dilution of the message, prompting consumers to stick to the status quo. An alternative explanation is that consumers were less willing to shop around as they may not have understood the gains they could make.



Figure 10: Rates of shopping around by treatment

#### Conclusions

The experimental results provide additional evidence that carefully designed information prompts can have an impact on consumer behaviour and encourage shopping around. Importantly, the experiment identified which information prompts resulted in more shopping around.

## 6 Lessons learned

Over the last three years, we have learned a number of lessons about carrying out behavioural research in financial regulation, both from our own and others' work. We have combined these practical lessons with insights from theory to develop the following list of tips which researchers may wish to consider when planning behavioural experiments.

#### 1. Diagnose the problem before trialling

It is crucial to diagnose and understand the problem before trialling potential remedies. Narrowing down too quickly, for example starting by testing the effectiveness of different types of disclosure before understanding whether disclosure is the most suitable solution, can mean that interventions with more impact will be missed.

While a narrow research focus is necessary and important when trialling (and can be particularly critical at the start of building research capability when buy-in from partners is all-important), there may be more that could be done. For example, in Experiment 1, *Encouraging compliance*, we investigated the potential causes of firms' non-submission of returns by talking to supervisors and looking at the forms to be submitted. However, given more time and resource, we might have benefitted from more qualitative research methods, such as interviewing the people in mutual societies responsible for submitting their forms. This could have helped us to understand the range of interventions which might improve compliance, whether this was the wording or timing of letter reminders or something more radical, such as changing the process for submission.

#### 2. Choose the right method

It is important to choose the right method and combine methods where possible when the situation is complex and/or under-researched. We found that lab experiments were very useful for questions where field trials were unfeasible, undesirable or at an earlier stage in the process of diagnosing problems and identifying potential solutions. For example, a lab experiment was a good method for investigating proposed standards for price comparison websites (see Experiment 7, *How much?*), largely because no such price comparison websites existed at the time! A lab experiment also allowed us to measure consumers' ability to select the cheapest deal in the absence of confounding factors (e.g. consumer preferences) that would be present in a field trial.

Another consideration in choosing the right method is the feasibility of trialling. Trials often take significant time and technical expertise to set up, implement and analyse and in some cases, we may need evidence to inform the policymaking process more quickly than this. In these cases, we may choose to use other research methods including lab experiments and quantitative and qualitative research. Related to this, as regulators, we currently ask firms to volunteer to take part in field trials with us; indeed, there are a number of routes for firms who would like to test with our support, including **Project Innovate's Sandbox**. However, this does mean that we may be limited in the interventions we test by firms' willingness to take part. Again, we may choose other methods in circumstances where agreement would be hard to reach in the available timeframe.

Sometimes it is less obvious which method would be most suitable and in these cases, decisions may be made on logistics and practical considerations, such as time, budget

and resource. For example, it would have been theoretically possible to run the annuity quote experiment described in *Shop and Save,* as a field trial. In this case, we chose a lab experiment because it meant we could test treatments more quickly and at lower cost. Furthermore, our prior work on diagnosing the problem did not show that customers were facing inertia, which meant that a lab experiment would be relatively similar to the real life environment and therefore more likely to have external validity.

#### 3. Where possible, test treatment design before going into the field

Assuming a field trial is the most appropriate method of research to investigate a particular question, unintended effects from treatments may be uncovered by subjecting potential treatments to lab experiments and/or qualitative research before going into the field. For example, in Experiment 4, *Please take your cash*, we found that a simple version of the letter actually increased calls to the firm, possibly because customers thought the letter was more likely to be a scam. Showing letters to a sample of potential customers either through a laboratory experiment or a depth interview might have identified this possibility, although in this case, resource constraints prevented this. We now aim to road test treatments for field trials in the lab or through qualitative work, before putting them in the field. We are also starting to experiment with more creative methods for designing trial treatments, including insights and methods from the fields of psychology, design and user-experience. This can help us maximise the effectiveness of our treatments and avoid them failing because of design faults.

Another factor to consider when designing treatments is how many attributes they vary on. In Experiment 2, *Help is here*, we tested versions of an email subject line which varied on more than one attribute. This was because the priority of the study was demonstrating 'proof of concept' of A/B testing, rather than theoretical precision. Consequently, it was not possible to determine which precise change had caused an increase in open rates. In an ideal world, it is preferable to systematically test changes informed by behavioural theory one at a time, unless there are very good (practical) reasons not to do so.

#### 4. Choose a realistic but sufficient sample size to detect important effects

Usually, trial planning should include power calculations to determine the appropriate sample size. Our handwritten envelopes trial (see Experiment 5, *Just for you*) was underpowered for logistical reasons and hence may not have been robust enough to detect effects, resulting in non-significant results. There may be good reasons to carry out experiments even when the sample size is underpowered, especially if the experiment is simple and low cost to carry out, but it is important to be aware that results may be inconclusive and to factor this into decisions about allocation of resources to experiments.

It is also important to be aware that power calculations require knowledge about the likely effect size, usually taken from existing research, and this can be distorted due to publication bias. One way to overcome this is to take the smallest effect the researcher would be interested in (for example, based on the cost of the treatment, the implementing organisation would be interested in an effect which helped X additional people, or generating £X extra redress) and then to use this value in the calculation to determine sample size (Simonsohn, 2014). This is a pragmatic approach, which means that real but small effects might be missed, but has the advantage that results are sure to be of practical interest.

#### 5. Plan post-trial analysis

Planning post-trial analysis in advance helps to keep questions focused and avoid "fishing" for significant or favourable results. In particular, it is important to choose the right outcome measures and agree them with implementing partners. For example, in Experiment 4, *Please take your cash*, we measured both the proportion of redress claims (primary measure) and the number of queries the firm received (secondary measure) for each treatment. This allowed us to see any unexpected or unintended effects of our intervention, as well as the effect on ultimate redress. However, using multiple outcome measures gives more freedom to selectively report significant or more interesting findings, which may lead to distortion of the findings. It is sensible to specify both primary and secondary measures in advance as far as possible, while remaining flexible and adaptive if experiments do not go as planned.

#### 6. Learn from null and negative results

While it may be disappointing to discover that interventions did not work as expected, this is important information, and can help us to avoid spending time, effort and money on activities which have a negligible or no effect. This can often be down to context and further testing can help us to understand the contexts in which certain interventions do and do not work so well. Sometimes we may see unexpected or even adverse effects of our interventions, and again, getting evidence of this can help us to adjust our activities and prevent potential harm.

Communicating null results can be a challenge. It is therefore important to manage the expectations of partners and stakeholders in advance of trialling. This allows us to ensure that all research is useful, since it is largely through failure that we can learn, adapt and expand our knowledge.

## References

Adams, P. and Hunt, S. (2013). Encouraging consumers to claim redress; evidence from a field trial. *Financial Conduct Authority, Occasional Paper 2.* 

Adams, P., Hunt, S., Vale, L., Zaliauskas, R. (2015). Stimulating interest: reminding savers to act when rates decrease. *Financial Conduct Authority, Occasional Paper 7.* 

Adams, P., Baker, R., Hunt, S., Kelly, D., Nava, A. (2015). Encouraging consumers to act at renewal: evidence from field trials in the home and motor insurance markets. *Financial Conduct Authority, Occasional Paper 12.* 

Adams, P., Hunt, S., Palmer, C., Zaliauskas, R. (2016). Attention, search and switching: evidence on mandated disclosure from the savings market. *Financial Conduct Authority, Occasional Paper 19.* 

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines and performance: self-control by precommitment . *Psychological Science, 13*(3), 2119-224.

Behavioural Insights Team (2014). EAST: Four simple ways to apply behavioural insights.

Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J.A., Perugini, M., Spies, J.R., & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217-224.

Brown, J. R., Kling, J. R., Mullainathan, S., & Wrobel, V. M. (2008). Why Don't People Insure Late-Life Consumption? A Framing Explanation of the Under-Annuitization Puzzle. *American Economic Review*, *98*(2): 304-09.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *PubMed*, 145-53.

Deaton, A. & Cartwright, N. (2016). Understanding and misunderstanding randomised controlled trials. No. w22595. *National Bureau of Economic Research*, 2016.

Fellner, G., Sausgruber, R., & Traxler, C. (2013). Testing enforcement strategies in the field: threat, moral appeal and social information. *Journal of the European Economic Association*, *11*(3), 634 – 660.

Financial Conduct Authority. (2014a). Does the framing of retirement income options matter?

Financial Conduct Authority. (2014b). Retirement income market study: interim report – provisional findings and proposed remedies.

Financial Conduct Authority. (2016). Retirement Income Market Data: July – September 2015.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502-1505.

Hunt, S., Stewart, N., Zaliauskas, R. (2015). Two plus two makes five? Survey evidence that investors overvalue structured deposits. *Financial Conduct Authority, Occasional Paper 9.* 

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS MEDICINE*, 2(8), e124.

Iscenko, Z., Duke, C., Huck, S., Wallace, B. (2014). How does selling insurance as an add-on affect consumer decisions? A practical application of behavioural experiments in financial regulation. *Financial Conduct Authority, Occasional Paper 3.* 

Irish Revenue (2013). Survey of SME Taxpayers 2013.

Karlan, D., & Appel, J. (2016). *Failing in the field: what we can learn when field research goes wrong.* Princeton University Press.

Kessler, J. B. & Vesterlund, L. (2012). The external validity of laboratory experiments: qualitative rather than quantitative effects. In Fréchette, Guillaume and Andrew Schotter (editors) *The Methods of Modern Experimental Economics*, Oxford University Press.

Koelle, F., Lane, T., Nosenzo, D., & Starmer, C. (2016). Nudging the electorate: what works and why? *Paper in preparation.* 

London Economics & YouGov. (2015). High-cost short-term credit price comparison websites: a behavioural study for the Financial Conduct Authority.

Mathieu, S., Boutron, I., Moher, D., Altman, D. G., & Ravaud, P. (2009). Comparison of registered and published primary outcomes in randomized controlled trials. *The Journal of the American Medical Association*, *302*(9), 977-984.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science 349*(6251).

Optimisa Research (2013), 'The Annuity Purchasing Process', Qualitative research report, prepared for the Financial Services Consumer Panel, 3–4.

Oxera and the Nuffield Centre for Experimental Social Sciences. (2016). Increasing consumer engagement in the annuities market: can prompts raise shopping around?

Rosenthal, R. (1979). The "File Drawer Problem" and Tolerance for Null Results. *Psychological Bulletin, 86*(3), 638-641.

Rothstein, H. R., Sutton, Alexander. J., Borenstein, M. (2005). Publication Bias in Meta-Analysis. In Rothstein, H. R., Sutton, Alexander. J., Borenstein, M. (Eds.). *Publication Bias in Meta-Analysis – Prevention, Assessment and Adjustments*. John Wiley & Sons, Ltd, Chichester, UK.

Schooler, J. W. (2011). Unpublished results hide the decline effect. Nature, 470, 437.

Sedlmeier, P. & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*(2), 309 – 316.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366.

Simonsohn, U. (2014). *Go big or go home: A practical alternative to impractical power analysis.* Talk presented at the annual conference of the Society of Personality and Social Psychology, Austin, TX.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician, 49,* 108–112.

Tversky, A. & Kahneman, D (1971). Belief in the Law of Small Numbers. *Psychological Bulletin,* 76(2), 105-110.

Wu, F. & Huberman, B. A. (2008). Popularity, novelty and attention. *Proceedings of the 9<sup>th</sup> ACM conference on Electronic Commerce*, 240-245.

© Financial Conduct Authority 2016 25 The North Colonnade Canary Wharf London E14 5HS Telephone: +44 (0)20 7066 1000 Website: www.fca.org.uk All rights reserved